Lecture 9

OLS can be used to estimate non-linear relationships, by transforming a non-linear form into a linear one

- Testing Functional Form

How do you know whether to use logs or levels for the dependent variable?

- Functional form : test of normality

(validity of t and F tests hang on assumption about normality in residuals)

- Multiple regression analysis

(look to see whether estimation methods and all the tests done so far carry over to case where there is more than 1 explanatory (X) variable

Testing Functional Form
How do you know whether to use logs or levels for the dependent variable?

If want to compare goodness of fit of models in which the dependent variable is in logs or levels then can not use the R^2.

The TSS in Y is not the same as the TSS in LnY,

$$\sum_{i=1}^{N}(Y_i - \bar{Y})^2 \neq \sum_{i=1}^{N}(LnY_i - \overline{LnY})^2$$

so comparing R^2 is not valid.

Instead the basic idea behind testing for the appropriate functional form of the *dependent* variable is to transform the data so as to make the ***RSS*** comparable

Do this by :

Do this by :

1. Calculating the geometric mean

where geometric (rather than arithmetic) mean

$$= (y_1 * y_2 * \ldots y_n)^{1/n} = \exp^{1/n \mathrm{Ln}(y_1 * y_2 \ldots y_n)}$$

Do this by :

1. Calculating the geometric mean

where geometric (rather than arithmetic) mean

$$= (y_1 * y_2 * \ldots y_n)^{1/n} = \exp^{1/n \mathrm{Ln}(y_1 * y_2 \ldots y_n)}$$

2. rescale each y observation by dividing by this value

$$y_i^* = y_i \text{ /geometric mean}$$

Do this by :

1. Calculating the geometric mean

where geometric (rather than arithmetic) mean

$$= (y_1 * y_2 * \ldots y_n)^{1/n} = \exp^{1/n Ln(y_1 * y_2 \ldots y_n)}$$

2. rescale each y observation by dividing by this value

$$y_i^* = y_i \text{ /geometric mean}$$

3. regress y^* (rather than y) on X, save RSS
   regress Lny^* (rather than Lny) on X, save RSS

Do this by :

1. Calculate the geometric mean

where geometric (rather than arithmetic) mean

$$= (y_1 * y_2 * \ldots y_n)^{1/n} = \exp^{1/n Ln(y_1 * y_2 \ldots y_n)}$$

2. rescale each y observation by dividing by this value

$$y_i^* = y_i \text{ /geometric mean}$$

3. regress y^* (rather than y) on X, save RSS
regress Lny^* (rather than Lny) on X, save RSS
(in practice the RSS is the same whether you use LnY or Lny^*)

the model with the lowest RSS is the one with the better fit

More formally can show that

$$BoxCox = N/2*\log(RSS_{largest}/RSS_{smallest}) \sim \chi^2_{(1)}$$

More formally can show that

$$BoxCox = N/2*\log(RSS_{largest}/RSS_{smallest}) \sim \chi^2_{(1)}$$

Follows a Chi-Squared distribution with one degree of freedom (one because there is one statistic being tested)

More formally can show that

BoxCox = $N/2 * \log(RSS_{largest}/RSS_{smallest}) \sim \chi^2_{(1)}$

Follows a Chi-Squared distribution with one degree of freedom (one because there is one statistic being tested)

If estimated value exceeds critical value (from tables Chi-squared at 5% level with 1 degree of freedom is 3.84) *reject* the null hypothesis that the models are the same

(ie there is a significantly different in terms of goodness of fit).

More formally can show that

$$BoxCox = N/2 * \log(RSS_{largest}/RSS_{smallest}) \sim \chi^2_{(1)}$$

Follows a Chi-Squared distribution with one degree of freedom (one because there is one statistic being tested)

If estimated value exceeds critical value (from tables Chi-squared at 5% level with 1 degree of freedom is 3.84) *reject* the null hypothesis that the models are the same

(ie there is a significantly different in terms of goodness of fit).

So choose the one with the ***lowest*** RSS

More formally can show that

$$BoxCox = N/2*\log(RSS_{largest}/RSS_{smallest}) \sim \chi^2_{(1)}$$

Follows a Chi-Squared distribution with one degree of freedom (one because there is one statistic being tested)

If estimated value exceeds critical value (from tables Chi-squared at 5% level with 1 degree of freedom is 3.84) *reject* the null hypothesis that the models are the same

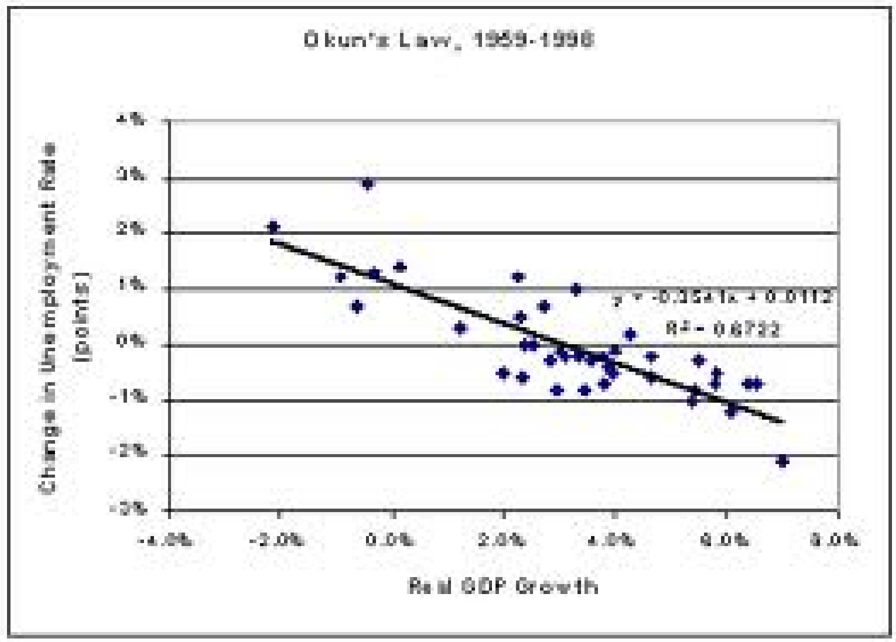
(ie there is a significantly different in terms of goodness of fit).

So choose the one with the ***lowest*** RSS

But do **not** use the transformed model to look at the coefficients of the model – use the originals

Example: Okun’s Law

Arthur Okun 1928-80 is an observed correlation between GDP growth and the unemployment or employment rates



```
. u boxcox                          /* read in data */
```

The data contains info on GDP and employment growth for 21 countries

```
. su empl gdp
    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        empl |      21    1.108095   .8418647        .02       3.02
         gdp |      21    3.059524   1.625172       1.15       7.73
```

The data show that gdp and employment growth are measured in percentage points, with a maximum of 7.73 %point annual GDP growth and a minimum 1.15% points.

A linear regression gives

```
. reg empl gdp
  Source |       SS       df       MS                  Number of obs =      21
---------+------------------------------               F(  1,    19) =   26.97
   Model |  8.31618159     1  8.31618159               Prob > F      =  0.0001
Residual |  5.85854191    19  .308344311               R-squared     =  0.5867
---------+------------------------------               Adj R-squared =  0.5649
   Total |  14.1747235    20  .708736175               Root MSE      =  .55529
------------------------------------------------------------------------------
    empl |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gdp |    .396778   .0764018      5.193   0.000       .2368672    .5566888
   _cons |  -.1058566   .2632937     -0.402   0.692      -.6569367    .4452235
```

Gdp is measured in percentage points, dempl/dgdp = β_{gdp}
and hence dempl= β_{gdp}* dgdp so a **1 % point** rise in gdp growth raises employment growth by 0.4 points a year

and a log-lin specification gives

```
g lempl=log(empl)                         /* generate log of dep. Variable */

. reg lempl gdp
  Source |       SS       df       MS                  Number of obs =      21
---------+------------------------------               F(  1,    19) =    5.89
   Model |  6.84252682     1  6.84252682               Prob > F      =  0.0253
Residual |  22.0706507    19   1.1616132               R-squared     =  0.2367
---------+------------------------------               Adj R-squared =  0.1965
   Total |  28.9131775    20  1.44565888               Root MSE      =  1.0778
------------------------------------------------------------------------------
   lempl |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gdp |     .35991   .1482915      2.427   0.025       .0495322    .6702877
   _cons |  -1.436343   .5110381     -2.811   0.011      -2.505958   -.3667282
```

log-lin model so coefficients are growth rates. This time dlempl/dgdp = β_{gdp}
and hence dlempl= β_{gdp}* dgdp where dlempl= % change in gdp/100.
So a 1% **point** (**not a 1 %)** rise in gdp growth raises emp growth by 36% a year (from table of means above, can see a 35% increase in gdp amounts to around 0.36 percentage points of extra growth a year – which is similar to estimate in levels)

Looks like linear specification is preferred, but since R^2 or RSS not comparable use Box-Cox test to test formally

Get geometric mean

```
. means empl
```

```
Variable |     Type        Obs        Mean       [95% Conf. Interval]
---------+-----------------------------------------------------------
    empl | Arithmetic      21     1.108095         .724883    1.491307
         |  Geometric      21     .7152021         .413749    1.236291
```

Rescale linear dependent variable and log of dependent variable

```
. g empadj=empl/.715
. g lempadj=log(empadj)
```

Regress adjusted dependent variables on gdp and log(gdp) respectively

```
. reg empadj gdp

  Source |       SS       df       MS              Number of obs =      21
---------+------------------------------           F(  1,    19) =   26.97
   Model |  16.2671653     1  16.2671653           Prob > F      =  0.0001
Residual |  11.4598119    19  .603147995           R-squared     =  0.5867
---------+------------------------------           Adj R-squared =  0.5649
   Total |  27.7269772    20  1.38634886           Root MSE      =  .77663

------------------------------------------------------------------------------
  empadj |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gdp |   .5549343   .1068557      5.193   0.000       .3312828    .7785858
   _cons |  -.1480511    .368243     -0.402   0.692      -.9187925    .6226903

. reg lempadj gdp

  Source |       SS       df       MS              Number of obs =      21
---------+------------------------------           F(  1,    19) =    5.89
   Model |  6.84252671     1  6.84252671           Prob > F      =  0.0253
Residual |  22.0706501    19  1.16161317           R-squared     =  0.2367
---------+------------------------------           Adj R-squared =  0.1965
   Total |  28.9131769    20  1.44565884           Root MSE      =  1.0778

------------------------------------------------------------------------------
 lempadj |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     gdp |     .35991   .1482915      2.427   0.025       .0495322    .6702877
   _cons |  -1.100871   .5110381     -2.154   0.044      -2.170486   -.0312554
```

Now RSS are comparable, and can see linear is preferred.

Formal test of significant difference between the 2 specifications

. g test=(21/2)*log(22.1/11.5) $= N/2\log(RSS_{largest}/RSS_{smallest}) \sim \chi^2_{(1)}$

/* stata recognises "log" as Ln or $\log_e$ */

```
. di test

  6.86
```

Given test is Chi-Squared with 1 degree of freedom. Estimated value exceeds critical value (from tables Chi-squared at 5% level with 1 degree of freedom is 3.84) so models are significantly different in terms of goodness of fit.

Test for Normality of Residuals

All the hypotheses, tests and confidence intervals done so far are based on the assumption that the (unknown true) residuals are normally distributed. If not then tests are invalid

Test for Normality of Residuals

All the hypotheses, tests and confidence intervals done so far are based on the assumption that the (unknown true) residuals are normally distributed. If not then tests are invalid

When choosing a functional form better to choose one which gives normally distributed errors

Since never observe true residuals can instead look at the *OLS residuals*

Test for Normality of Residuals

All the hypotheses, tests and confidence intervals done so far are based on the assumption that the (unknown true) residuals are normally distributed. If not then tests are invalid

When choosing a functional form better to choose one which gives normally distributed errors

Since never observe true residuals can instead look at the *OLS residuals*

Why?

Test for Normality of Residuals

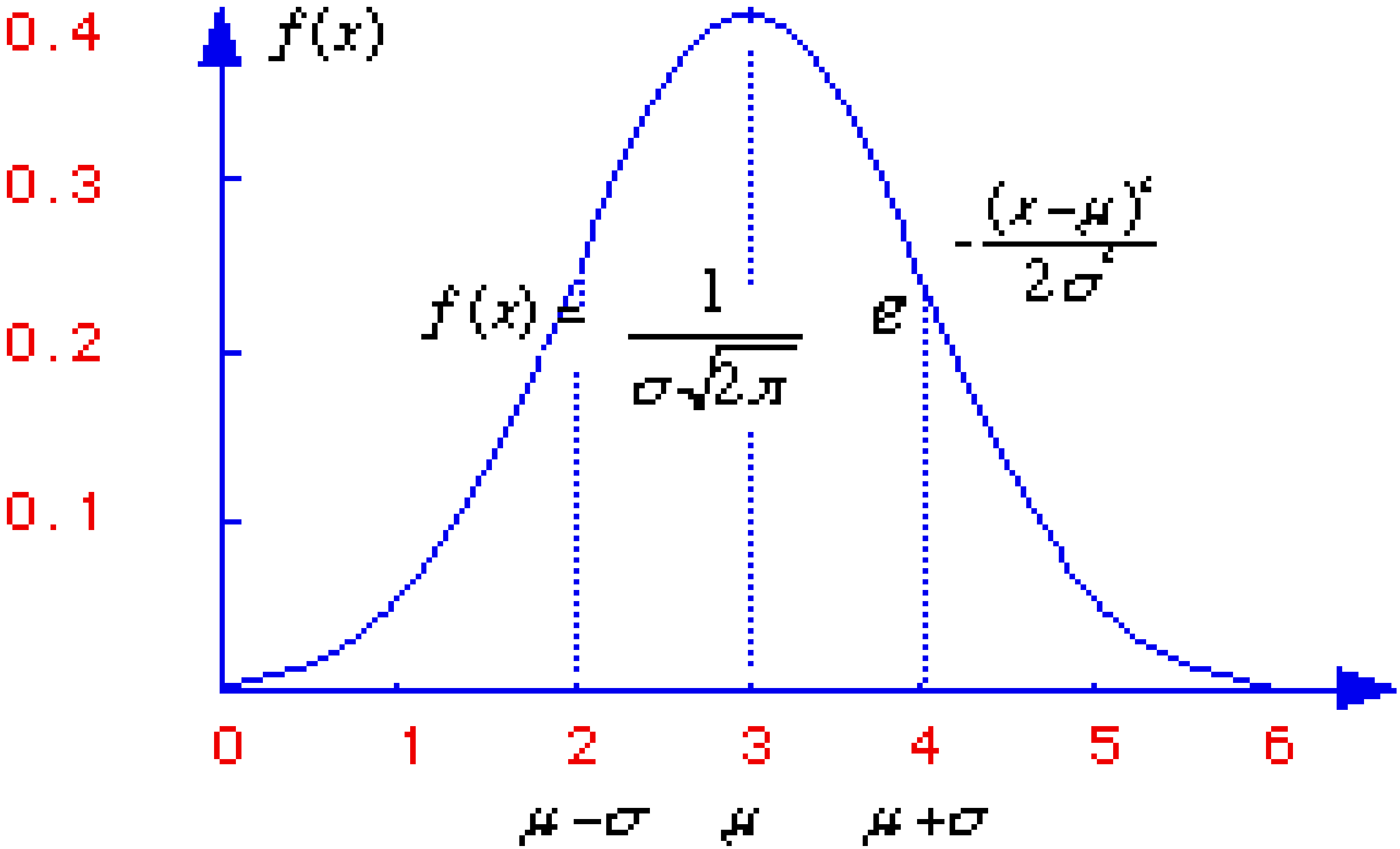
All the hypotheses, tests and confidence intervals done so far are based on the assumption that the (unknown true) residuals are normally distributed. If not then tests are invalid

When choosing a functional form better to choose one which gives normally distributed errors

Since never observe true residuals can instead look at the *OLS residuals*

Why? Can show that *if* all Gauss-Markov assumptions are satisfied (see earlier notes) then the OLS residuals are also *asymptotically* normally distributed (ie approximately normal if sample size is large)

A normal distribution should have following properties:



A normal distribution should have following properties:

1. symmetric about its mean (in the case of OLS residuals the mean is zero)

A normal distribution should have following properties:

1. symmetric about its mean (in the case of OLS residuals the mean is zero)

A Non-symmetric distribution is said to be *skewed*.

A normal distribution should have following properties:

1. symmetric about its mean (in the case of OLS residuals the mean is zero)

A Non-symmetric distribution is said to be *skewed*.

Can measure this by looking at the 3^{rd} moment of the normal distribution relative to the 2^{nd} (mean is the 1^{st} moment, variance is the second moment)

$$Skewness = \frac{[E(X-\mu_X)^3]^2}{[E(X-\mu_X)^2]^3} = \frac{square of\ 3^{rd}\ moment}{cube of\ 2^{nd}\ moment}$$

A normal distribution should have following properties:

1. symmetric about its mean (in the case of OLS residuals the mean is zero)

A Non-symmetric distribution is said to be *skewed*.

Can measure this by looking at the 3rd moment of the normal distribution relative to the 2nd (mean is the 1st moment, variance is the second moment)

$$Skewness = \frac{[E(X-\mu_X)^3]^2}{[E(X-\mu_X)^2]^3} = \frac{square of\ 3^{rd}\ moment}{cube of\ 2^{nd}\ moment}$$

Symmetry is represented by a value of 0 for the skewness coefficient

A normal distribution should have following properties:

1. symmetric about its mean (in the case of OLS residuals the mean is zero)

A Non-symmetric distribution is said to be *skewed*.

Can measure this by looking at the 3rd moment of the normal distribution relative to the 2nd (mean is the 1st moment, variance is the second moment)

$$Skewness = \frac{[E(X-\mu_X)^3]^2}{[E(X-\mu_X)^2]^3} = \frac{square of\ 3^{rd}\ moment}{cube of\ 2^{nd}\ moment}$$

Symmetry is represented by a value of 0 for the skewness coefficient

Right skewness gives a value > 0
(more values clustered to close to left of mean and a few values a long way to the right of the mean tend to make the value >0)

A normal distribution should have following properties:

1. symmetric about its mean (in the case of OLS residuals the mean is zero)

A Non-symmetric distribution is said to be *skewed*.
Can measure this by looking at the 3rd moment of the normal distribution relative to the 2nd (mean is the 1st moment, variance is the second moment)

$$Skewness = \frac{[E(X-\mu_X)^3]^2}{[E(X-\mu_X)^2]^3} = \frac{square of\ 3^{rd}\ moment}{cube of\ 2^{nd}\ moment}$$

Symmetry is represented by a value of zero for the skewness coefficient

Right skewness gives a value > 0
(more values clustered to close to left of mean and a few values a long way to the right of the mean tend to make the value >0)

Left skewness gives a value < 0

(kdensity age, normal)

2. A distribution is said to display ***kurtosis*** if the height of the distribution is unusual

2. A distribution is said to display ***kurtosis*** if the height of the distribution is unusual

(suggests observations more bunched or more spread out than should be).

2. A distribution is said to display ***kurtosis*** if the height of the distribution is unusual

(suggests observations more bunched or more spread out than should be).

Measure this by

$$Kurtosis = \frac{E(X-\mu_X)^4}{[E(X-\mu_X)^2]^2} = \frac{4^{th}\, moment}{square of\ 2^{nd}\, moment}$$

2. A distribution is said to display ***kurtosis*** if the height of the distribution is unusual

(suggests observations more bunched or more spread out than should be).

Measure this by

$$Kurtosis = \frac{E(X-\mu_X)^4}{[E(X-\mu_X)^2]^2} = \frac{4^{th}\,moment}{square of\ 2^{nd}\,moment}$$

A normal distribution should have a kurtosis value of 3

Can combine both these features to give the **Jarque-Bera Test for Normality** (in the residuals)

$$JB = N * \left[\frac{Skewness^2}{6} + \frac{(Kurtosis - 3)^2}{24} \right]$$

Can combine both these features to give the **Jarque-Bera Test for Normality** (in the OLS residuals, since true residuals unobserved)

$$JB = N * \left[\frac{Skewness^2}{6} + \frac{(Kurtosis - 3)^2}{24} \right]$$

Can show that this is *asymptotically* Chi^2 distributed with 2 degrees of freedom (1 for skewness and 1 for kurtosis)

Can combine both these features to give the **Jarque-Bera Test for Normality** (in the OLS residuals, since true residuals unobserved)

$$JB = N * \left[\frac{Skewness^2}{6} + \frac{(Kurtosis - 3)^2}{24} \right]$$

Can show that this is *asymptotically* Chi^2 distributed with 2 degrees of freedom (1 for skewness and 1 for kurtosis)

If estimated chi-squared > $\text{chi-squared}_{critical}$

reject null that residuals are normally distributed

Can combine both these features to give the **Jarque-Bera Test for Normality** (in the OLS residuals, since true residuals unobserved)

$$JB = N * \left[\frac{Skewness^2}{6} + \frac{(Kurtosis - 3)^2}{24} \right]$$

Can show that this is *asymptotically* Chi^2 distributed with 2 degrees of freedom (1 for skewness and 1 for kurtosis)

If estimated chi-squared > $\text{chi-squared}_{critical}$

reject null that residuals are normally distributed

(If not suggests should try another functional form to try and make residuals normal, otherwise t stats may be invalid).

Example: **Jarque-Bera Test for Normality (in residuals)**

```
. u wage                            /* read in data */
```

1st regress hourly pay on years of experience and get residuals

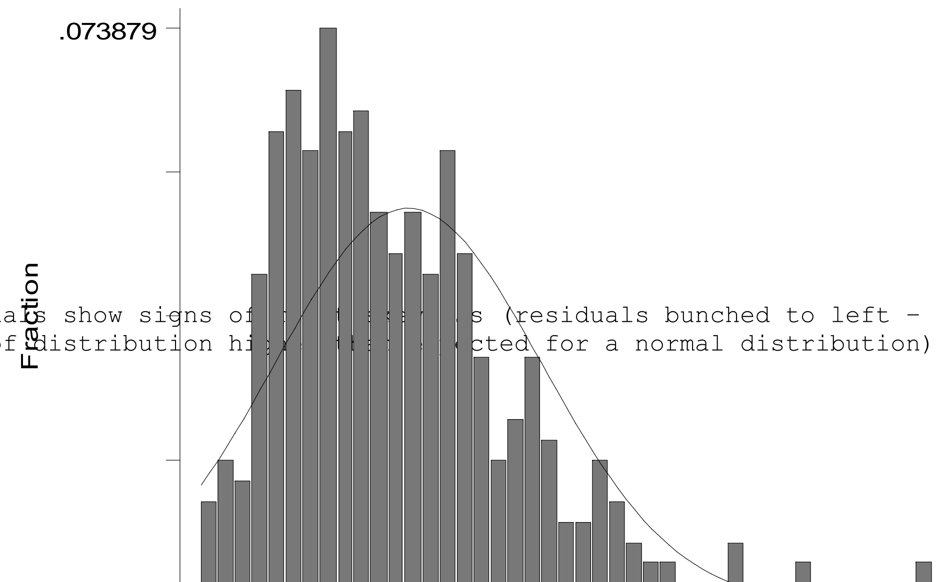
```
. reg hourpay xper

  Source |       SS       df       MS                  Number of obs =     379
---------+------------------------------               F(  1,   377) =    7.53
   Model |  136.061219     1  136.061219               Prob > F      =  0.0064
Residual |  6815.41926   377  18.0780352               R-squared     =  0.0196
---------+------------------------------               Adj R-squared =  0.0170
   Total |  6951.48048   378    18.39016               Root MSE      =  4.2518
------------------------------------------------------------------------------
 hourpay |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    xper |   .0487259    .017761      2.743   0.006       .0138028     .083649
   _cons |    7.26455   .4333534     16.764   0.000       6.412457    8.116642
------------------------------------------------------------------------------

. predict res, resid
```

Check histogram of residuals using the following stata command

```
. gra res, normal bin(50)
      /* normal option superimposes a normal distribution on the graph */
```



Residuals show signs of skewness (residuals bunched to left – not symmetric) and kurtosis (leptokurtic – since peak of distribution higher than expected for a normal distribution)

```
To test more formally

. su res, detail

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -6.253362      -6.580268
 5%    -4.919813      -6.372607
10%     -4.27017      -6.313276       Obs                 379
25%    -3.011451      -6.253362       Sum of Wgt.         379

50%    -.9261839                      Mean           1.11e-08
                        Largest       Std. Dev.      4.246199
75%     1.869452        16.5097
90%     5.383683       17.73377       Variance       18.03021
95%     7.480312        17.9211       Skewness        1.50555
99%      16.5097       20.44043       Kurtosis       6.432967
```

Construct Jarque-Bera test

. jb = (379/6)((1.50555^2)+(((6.43-3)^2)/4))*

= 328.9

*The statistic has a Chi^2 distribution with **2** degrees of freedom, (one for skewness one for kurtosis).*

From tables critical value at 5% level for 2 degrees of freedom is 5.99

*So JB>$\chi^2_{critical}$, so **reject** null that residuals are normally distributed.*

Suggests should try another functional form to try and make residuals normal, otherwise t stats may be invalid.

Remember this test is only valid asymptotically, so it relies on having a large sample size. Users with data sets smaller than 50 observations should be wary about using this test.

N.B. Stata can do this automatically if you download the"jb6" command
Just type "ssc install jb6" to install this command

jb6 res
Jarque-Bera normality test: 329.3 Chi(2) 3.1e-72
Jarque-Bera test for Ho: normality: (uhat)

Multiple Regression Analysis

In most cases unlikely can explain all of behaviour in the dependent variable by a single explanatory variable. Most problems require 2 or more right hand side variables to capture behaviour adequately.

Multiple Regression Analysis

In most cases unlikely can explain all of behaviour in the dependent variable by a single explanatory variable. Most problems require 2 or more right hand side variables to capture behaviour adequately.

Consider case of two explanatory variables:

Suppose for example that rather than model wages as depending on a single variable, age

$$wage = \beta_0 + \beta_1 Age + u$$

Multiple Regression Analysis

In most cases unlikely can explain all of behaviour in the dependent variable by a single explanatory variable. Most problems require 2 or more right hand side variables to capture behaviour adequately.

Consider case of two explanatory variables:

Suppose for example that rather than model wages as depending on a single variable, age

$$wage = \beta_0 + \beta_1 Age + u$$

We let wages depend on 2 variables: age and education

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

Multiple Regression Analysis

In most cases unlikely can explain all of behaviour in the dependent variable by a single explanatory variable. Most problems require 2 or more right hand side variables to capture behaviour adequately.

Consider case of two explanatory variables:

Suppose for example that rather than model wages as depending on a single variable, age

$$wage = \beta_0 + \beta_1 Age + u$$

We let wages depend on 2 variables: age and education

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

Economic intuition (theory) tells us that wages should increase with age and also increase with number of years of schooling (both thought to raise marginal productivity through different channels (experience and skill) so would like to estimate their separate effects.

Multiple Regression Analysis

In most cases unlikely can explain all of behaviour in the dependent variable by a single explanatory variable. Most problems require 2 or more right hand side variables to capture behaviour adequately.

Consider case of two explanatory variables:

Suppose for example that rather than model wages as depending on a single variable, age

$$wage = \beta_0 + \beta_1 Age + u$$

We let wages depend on 2 variables: age and education

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

Economic intuition (theory) tells us that wages should increase with age and also increase with number of years of schooling (both thought to raise marginal productivity through different channels (experience and skill) so would like to estimate their separate effects.

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

The interpretation of the coefficients now corresponds to the *ceteris paribus* (other things equal) assumption often made in economic theory,

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

The interpretation of the coefficients now corresponds to the *ceteris paribus* (other things equal) assumption often made in economic theory,

- the presence of schooling now “nets out” the influence on age rather than relying on its influence through the residuals as in 2 variable model

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

The interpretation of the coefficients now corresponds to the *ceteris paribus* (other things equal) assumption often made in economic theory,

- the presence of schooling now “nets out” the influence on age rather than relying on its influence through the residuals as in 2 variable model

ie before years of schooling is part of all the unmeasured stuff contained in the residual

$$wage = \beta_0 + \beta_1 Age + \{Yearsofschooling + otherstuff\}$$

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

The interpretation of the coefficients now corresponds to the *ceteris paribus* (other things equal) assumption often made in economic theory,

- the presence of schooling now "nets out" the influence on age rather than relying on its influence through the residuals as in 2 variable model

ie before years of schooling is part of all the unmeasured stuff contained in the residual

$$wage = \beta_0 + \beta_1 Age + \{Yearsofschooling + otherstuff\}$$

Now explicitly modelling its effect – so it is no longer part of the residual

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

The interpretation of the coefficients now corresponds to the *ceteris paribus* (other things equal) assumption often made in economic theory,

- the presence of schooling now "nets out" the influence on age rather than relying on its influence through the residuals as in 2 variable model

ie before years of schooling is part of all the unmeasured stuff contained in the residual

$$wage = \beta_0 + \beta_1 Age + \{Yearsofschooling + otherstuff\}$$

Now explicitly modelling its effect – so it is no longer part of the residual

$$wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$$

- This means that the estimated coefficient on age can now be considered as holding schooling constant (ie "other things equal")

More formally:

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides of (2))

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides)

We want to isolate the effect of age on the wage when schooling is held fixed

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides)

We want to isolate the effect of age on the wage when schooling is held fixed

Schooling fixed implies that

$\Delta Yearsofschooling = 0$ (no change)

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)
(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides)

We want to isolate the effect of age on the wage when schooling is held fixed

Schooling fixed implies that

ΔYearsofschooling=0 (no change)

So that in this case

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age$$

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides)

We want to isolate the effect of age on the wage when schooling is held fixed

Schooling fixed implies that

$\Delta Yearsofschooling = 0$ (no change)

So that in this case

$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age$ and hence $\Delta \hat{wage} / \Delta Age = \hat{\beta}_1$

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides)

We want to isolate the effect of age on the wage when schooling is held fixed

Schooling fixed implies that

$\Delta Yearsofschooling = 0$ (no change)

So that in this case

$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age$ and hence $\Delta \hat{wage} / \Delta Age = \hat{\beta}_1$

Hence multiple OLS regression coefficients are said to be equivalent to *partial derivatives* holding the effect of the other variables fixed (ie set to zero change)

Given the model $wage = \beta_0 + \beta_1 Age + \beta_2 Yearsofschooling + u$ (1)

the ols prediction is $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 Yearsofschooling$ (2)

(estimated values denoted by ^)

and it follows that change in the wage is

$$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age + \hat{\beta}_2 \Delta Yearsofschooling$$

(just difference both sides)

We want to isolate the effect of age on the wage when schooling is held fixed

Schooling fixed implies that

$\Delta Yearsofschooling=0$ (no change)

So that in this case

$\Delta \hat{wage} = \hat{\beta}_1 \Delta Age$ and hence $\Delta \hat{wage} / \Delta Age = \hat{\beta}_1$

Hence multiple OLS regression coefficients are said to be equivalent to *partial derivatives* holding the effect of the other variables fixed (ie set to zero change)

$$\left.\frac{\partial Y}{\partial X_1}\right|_{allotherX constant} \Rightarrow \left.\frac{\partial Wage}{\partial Age}\right|_{schooling constant}$$

Estimation of Multiple Regression by OLS

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So

$$RSS = \sum \hat{u}_i^2$$

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So

$$RSS = \sum \hat{u}_i^2 = \sum (Wage_i - \hat{Wage}_i)^2$$

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So $$RSS = \sum \hat{u}_i^2 = \sum \left(Wage_i - \hat{Wage}_i \right)^2 = \sum (Wage_i - \hat{\beta}_0 - \hat{\beta}_1 Age_i - \hat{\beta}_2 Schooling_i)^2$$

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So $RSS = \sum \hat{u}_i^2 = \sum \left(Wage_i - \hat{Wage}_i \right)^2 = \sum (Wage_i - \hat{\beta}_0 - \hat{\beta}_1 Age_i - \hat{\beta}_2 Schooling_i)^2$

or more generally

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 Age_i - \hat{\beta}_2 Schooling_i)^2$$

or more generally

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

The only difference between this and simple regression used before

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2$$

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 Age_i - \hat{\beta}_2 Schooling_i)^2$$

or more generally

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

The only difference between this and simple regression used before

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2$$

is that there are now 3 unknowns (β_0 β_1 and β_2) not 2 unknowns as before (β_0 β_1)

Estimation of Multiple Regression by OLS

The derivation of OLS coefficients is much as before. The idea remains to fit a straight line through the data and choose the coefficients that minimise the sum of squared residuals (the difference between the predictions from the line and the actual values)

(just harder to draw a straight line in 3-d or multiple dimensions)

In the example above the number of explanatory variables is 2, (Age & Schooling)

So

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 Age_i - \hat{\beta}_2 Schooling_i)^2$$

or more generally

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

The only difference between this and simple regression used before

$$RSS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2$$

is that there are now 3 unknowns (β_0 β_1 and β_2) not 2 unknowns as before (β_0 β_1) and 3 not 2 equations to solve for them (minimise above wrt β_0, β_1 and β_2)

The first order conditions for a minimum are

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0 = -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 = -2\sum_{i=1}^{N} X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 0 = -2\sum_{i=1}^{N} X_{2i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

The first order conditions for a minimum are

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0 = -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 = -2\sum_{i=1}^{N}X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 0 = -2\sum_{i=1}^{N}X_{2i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

Solving these 3 equations gives

The first order conditions for a minimum are

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0 = -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 = -2\sum_{i=1}^{N} X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 0 = -2\sum_{i=1}^{N} X_{2i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})$$

Solving these 3 equations gives

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1,Y)\text{Var}(X_2) - \text{Cov}(X_2,Y)\text{Cov}(X_1,X_2)}{\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1,X_2)]^2}$$

$$\hat{\beta}_2 = \frac{\text{Cov}(X_2,Y)\text{Var}(X_1) - \text{Cov}(X_1,Y)\text{Cov}(X_1,X_2)}{\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1,X_2)]^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)\text{Var}(X_2) - \text{Cov}(X_2, Y)\text{Cov}(X_1, X_2)}{\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1, X_2)]^2}$$

$$\hat{\beta}_2 = \frac{\text{Cov}(X_2, Y)\text{Var}(X_1) - \text{Cov}(X_1, Y)\text{Cov}(X_1, X_2)}{\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1, X_2)]^2}$$

The equations for the slope coefficients are similar to those in the 2 variable model, but contain extra terms which net out the influence of the other variables in explaining Y *and* the X variable of interest

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1,Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2,Y)\mathrm{Cov}(X_1,X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1,X_2)]^2}$$

$$\hat{\beta}_2 = \frac{\mathrm{Cov}(X_2,Y)\mathrm{Var}(X_1) - \mathrm{Cov}(X_1,Y)\mathrm{Cov}(X_1,X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1,X_2)]^2}$$

The equations are similar to those in the 2 variable model, but contain extra terms which net out the influence of the other variables in explaining Y *and* the x variable of interest

ie the difference in the OLS estimate of β_1 in the 2 variable model

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)}$$

ie the difference in the OLS estimate of β_1 in the 2 variable model

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1, Y)}{\mathrm{Var}(X_1)}$$

and the 3 variable model

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

ie the difference in the OLS estimate of β_1 in the 2 variable model

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)}$$

and the 3 variable model

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)\text{Var}(X_2) - \text{Cov}(X_2, Y)\text{Cov}(X_1, X_2)}{\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1, X_2)]^2}$$

depends on

a) the covariance between the variables, $\text{Cov}(X_1, X_2)$

ie the difference in the OLS estimate of β_1 in the 2 variable model

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1,Y)}{\mathrm{Var}(X_1)}$$

and the 3 variable model

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1,Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2,Y)\mathrm{Cov}(X_1,X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1,X_2)]^2}$$

depends on

a) the covariance between the variables, $\mathrm{Cov}(X_1, X_2)$

b) the influence of the omitted variable on the dependent variable, $\mathrm{Cov}(X_2,y)$

ie the difference in the OLS estimate of β_1 in the 2 variable model

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1, Y)}{\mathrm{Var}(X_1)}$$

and the 3 variable model

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1, Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, Y)\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)\mathrm{Var}(X_2) - [\mathrm{Cov}(X_1, X_2)]^2}$$

depends on

a) the covariance between the variables, $\mathrm{Cov}(X_1, X_2)$
b) the influence of the omitted variable on the dependent variable, $\mathrm{Cov}(X_2, y)$
c) the variance of the extra variable, $\mathrm{Var}(X_2)$

ie the difference in the OLS estimate of β$_1$ in the 2 variable model

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)}$$

And the 3 variable model

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)\text{Var}(X_2) - \text{Cov}(X_2, Y)\text{Cov}(X_1, X_2)}{\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1, X_2)]^2}$$

depends on

a) the covariance between the variables, Cov(X_1, X_2)

b) the influence of the omitted variable on the dependent variable, Cov(X_2,y)

c) the variance of the extra variable, Var(X_2)

so the multiple regression estimates “net out” the influence of other factors on both the dependent and explanatory variables

Properties of Multiple Regression Coefficients

Properties of Multiple Regression Coefficients

Can show that the properties of OLS estimators of the 2 variable model carry over into the general case,

Properties of Multiple Regression Coefficients

Can show that the properties of OLS estimators of the 2 variable model carry over into the general case, so that OLS estimators are always

Unbiased

Properties of Multiple Regression Coefficients

Can show that the properties of OLS estimators of the 2 variable model carry over into the general case, so that OLS estimators are always

- Unbiased
- Efficient (smallest variance of any unbiased estimator)

Properties of Multiple Regression Coefficients

Can show that the properties of OLS estimators of the 2 variable model carry over into the general case, so that OLS estimators are always

Unbiased
Efficient (smallest variance of any unbiased estimator)

With regard to the estimated standard errors on the coefficients in the 3 variable model can show that

With regard to the estimated standard errors on the coefficients in the 3 variable model can show that

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

With regard to the estimated standard errors on the coefficients in the 3 variable model can show that

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

where r^2_{x1x2} is the square of the correlation coefficient between X_1 & X_2

With regard to the estimated standard errors on the coefficients in the 3 variable model can show that

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

where r^2_{x1x2} is the square of the correlation coefficient between X_1 & X_2

(compared with $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)}$ in the 2 variable model)

The only other difference is that now the residual variance

The only other difference is that now the residual variance

$$s^2 = \frac{N * Var(\hat{u})}{N-k} = \frac{RSS}{N-k} = \frac{\sum_{i=1}^{N} \hat{u}^2}{N-k}$$

where k = no. of rhs coefficients (including the constant)

The only other difference is that now the residual variance

$$s^2 = \frac{N * Var(\hat{u})}{N-k} = \frac{RSS}{N-k} = \frac{\sum_{i=1}^{N} \hat{u}^2}{N-k}$$

where k = no. of rhs coefficients (including the constant)

(rather than $s^2 = \frac{N * Var(\hat{u})}{N-2} = \frac{RSS}{N-2} = \frac{\sum_{i=1}^{N} \hat{u}^2}{N-2}$

as in 2 variable model)

As before given $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$

As before given $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$

1) an increase in the residual variance, s^2

As before given $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$

an increase in the residual variance, s^2
a fall in sample size N

will make the OLS estimates of the effects of the X variables less precise

As before given $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$

an increase in the residual variance, s^2
a fall in sample size N

will make the OLS estimates of the effects of the X variables less precise

Now in addition
an increased correlation between X_1 & X_2 r^2_{x1x2}

will also make the OLS estimates of the effects of the X variables less precise (can't distinguish between the contribution of the individual variables if correlation is high)

Can show that the properties of OLS estimators of the 2 variable model carry over into the general case, so that OLS estimators are always

i) Unbiased
ii) Efficient (smallest variance of any unbiased estimator)

With regard to the estimated standard errors on the coefficients in the 3 variable model can show that

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

where r^2_{x1x2} is the square of the correlation coefficient between X_1 & X_2

(compared with $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)}$ in the 2 variable model)

As before given $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$

an increase in the residual variance, s^2
a fall in sample size N

will make the OLS estimates of the effects of the X variables less precise

Now in addition
an increased correlation between X_1 & X_2 r^2_{x1x2}

will also make the OLS estimates of the effects of the X variables less precise (can't distinguish between the contribution of the individual variables if correlation is high)

The consequences of this high correlation is called ***multicolinearity***

The consequences of this high correlation is called ***multicolinearity***

And the symptoms are that

1) while OLS estimates remain unbiased

The consequences of this high correlation is called ***multicolinearity***

And the symptoms are that

1) while OLS estimates remain unbiased
2) the standard errors are much larger than would be in the absence of multicolinearity

The consequences of this high correlation is called ***multicolinearity***

And the symptoms are that

while OLS estimates remain unbiased

the standard errors are much larger than would be in the absence of multicolinearity

and since $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$

the estimated t values will be smaller than otherwise.

The consequences of this high correlation is called ***multicolinearity***

And the symptoms are that

while OLS estimates remain unbiased

the standard errors are much larger than would be in the absence of multicolinearity

and since $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$

the estimated t values will be smaller than otherwise.

You may therefore conclude that variables are statistically insignificant (from zero) when not (ie Type II error)

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero,

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero, (if it is the variables are said to be orthogonal).

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero, (if it is the variables are said to be orthogonal).

The issue then becomes how serious a problem is it.

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero, (if it is the variables are said to be orthogonal).

The issue then becomes how serious a problem is it.

Detection:

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero, (if it is the variables are said to be orthogonal).

The issue then becomes how serious a problem is it.

Detection:

Low t values and high R^2

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero, (if it is the variables are said to be orthogonal).

The issue then becomes how serious a problem is it.

Detection:

Low t values and high R^2

The estimates may be sensitive to addition or subtraction of a small number of observations

In practice nearly all estimation suffers from multicolinearity since unlikely that the correlation between variables is zero, (if it is the variables are said to be orthogonal).

The issue then becomes how serious a problem is it.

Detection:

Low t values and high R^2

The estimates may be sensitive to addition or subtraction of a small number of observations

Look at the simple correlation coefficients between any 2 variables. A correlation coefficient >0.8 usually says there are problems.
(Or if the correlation between any two right hand side variables is greater than the correlation between that of each with the dependent variable)

(cost function eg)

Problem: In cases when there are many right hand side variables this strategy may not pick up *group* as opposed to *pairwise* correlations.

Problem: In cases when there are many right hand side variables this strategy may not pick up *group* as opposed to *pairwise* correlations.

In this case run an auxiliary regression of *any one* of the right hand side variables on *all* the other X variables

$$X_1 = \delta_0 + \delta_2 X_2 + \delta_3 X_3 + \ldots \delta_k X_k + u$$

and look at the R^2 from this regression (R_i^2).

Problem: In cases when there are many right hand side variables this strategy may not pick up *group* as opposed to *pairwise* correlations.

In this case run an auxiliary regression of *any one* of the right hand side variables on *all* the other X variables

$$X_1 = \delta_0 + \delta_2 X_2 + \delta_3 X_3 + \ldots \delta_k X_k + u$$

and look at the R^2 from this regression (R_i^2).

As a guide an $R_i^2 > 0.8$ suggests problems

In the general case (with k right hand side variables) it can be shown that

$$Var(\hat{\beta}_i) = \frac{s^2}{RSS_i} = \frac{s^2}{TSS_i * (1 - R_i^2)}$$

In the general case (with k right hand side variables) it can be shown that

$$Var(\hat{\beta}_i) = \frac{s^2}{RSS_i} = \frac{s^2}{TSS_i * (1 - R_i^2)}$$

where

RSS_i is the residual sum of squares from a regression of the variable X_i on all the other right hand side variables

In the general case (with k right hand side variables) it can be shown that

$$Var(\hat{\beta}_i) = \frac{s^2}{RSS_i} = \frac{s^2}{TSS_i * (1 - R_i^2)}$$

where

RSS_i is the residual sum of squares from a regression of the variable X_i on all the other right hand side variables

TSS_i is the total sum of squares from a regression of the variable X_i on all the other right hand side variables

In the general case (with k right hand side variables) it can be shown that

$$Var(\hat{\beta}_i) = \frac{s^2}{RSS_i} = \frac{s^2}{TSS_i * (1 - R_i^2)}$$

where

RSS_i is the residual sum of squares from a regression of the variable X_i on all the other right hand side variables

TSS_i is the total sum of squares from a regression of the variable X_i on all the other right hand side variables

R_i^2 is the R^2 from this regression

In the general case (with k right hand side variables) it can be shown that

$$Var(\hat{\beta}_i) = \frac{s^2}{RSS_i} = \frac{s^2}{TSS_i * (1 - R_i^2)}$$

where

RSS_i is the residual sum of squares from a regression of the variable X_i on all the other right hand side variables

TSS_i is the total sum of squares from a regression of the variable X_i on all the other right hand side variables

R_i^2 is the R^2 from this regression

Compared to $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$ in the 3 variable case

In the general case (with k right hand side variables) it can be shown that

$$Var(\hat{\beta}_i) = \frac{s^2}{RSS_i} = \frac{s^2}{TSS_i * (1 - R_i^2)}$$

where

RSS_i is the residual sum of squares from a regression of the variable X_i on all the other right hand side variables

TSS_i is the total sum of squares from a regression of the variable X_i on all the other right hand side variables

R_i^2 is the R^2 from this regression

Compared to $Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$ in the 3 variable case

Basic Point: The greater is this groupwise multi-colinearity, the larger the variance of (the less precise) the estimate

Example: Multicolinearity

Often in time series data when there are few observations (annual data is often all there is available) variables display common trends and so are highly correlated. This means it is difficult to discern individual effects of the RHS variables.

Suppose you regress consumption on a time trend, (a trend is just a variable that increases by one for each year of the data)

```
. reg cons trend

  Source |       SS       df       MS              Number of obs =      45
---------+------------------------------           F(  1,    43) =  960.81
   Model |  4.5380e+11     1  4.5380e+11           Prob > F      =  0.0000
Residual |  2.0309e+10    43   472306243           R-squared     =  0.9572
---------+------------------------------           Adj R-squared =  0.9562
   Total |  4.7411e+11    44  1.0775e+10           Root MSE      =   21733
------------------------------------------------------------------------------
    cons |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   trend |   7732.329   249.4543     30.997   0.000       7229.257    8235.402
   _cons |   129380.1   6588.931     19.636   0.000       116092.2    142667.9
------------------------------------------------------------------------------
```

This appears highly significant and economically important.

However a 3 variable regression of consumption on the trend and income gives

```
. reg cons trend income

  Source |       SS       df       MS              Number of obs =      45
---------+------------------------------           F(  2,    42) = 2919.99
   Model |  4.7072e+11     2  2.3536e+11           Prob > F      =  0.0000
Residual |  3.3853e+09    42  80603294.8           R-squared     =  0.9929
---------+------------------------------           Adj R-squared =  0.9925
   Total |  4.7411e+11    44  1.0775e+10           Root MSE      =  8977.9
------------------------------------------------------------------------------
    cons |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   trend |  -140.4874   553.0085     -0.254   0.801      -1256.504    975.5288
  income |   .9333721   .0644142     14.490   0.000       .8033789    1.063365
   _cons |   11579.25   8573.289      1.351   0.184      -5722.351    28880.84
------------------------------------------------------------------------------
```

The trend variable is now insignificant, the standard error on the estimate has increased massively and the sign of the coefficient is negative. This does not look sensible.

Suppose now drop just one observation from the data set

```
. reg cons trend income if year>55

  Source |       SS       df       MS              Number of obs =      44
---------+------------------------------           F(  2,    41) = 2746.58
   Model |  4.5073e+11     2  2.2536e+11           Prob > F      =  0.0000
Residual |  3.3641e+09    41  82052169.7           R-squared     =  0.9926
---------+------------------------------           Adj R-squared =  0.9922
   Total |  4.5409e+11    43  1.0560e+10           Root MSE      =  9058.3

------------------------------------------------------------------------------
    cons |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   trend |  -66.88367   576.4408     -0.116   0.908      -1231.029    1097.262
  income |    .926338   .0664476     13.941   0.000       .7921443    1.060532
   _cons |   12029.33   8695.204      1.383   0.174      -5530.987    29589.65
------------------------------------------------------------------------------
```

When we drop just one observation from the data the estimates again change noticeably.

Both these patterns are classic symptoms of multicolinearity. This can be confirmed by the simple pair-wise correlation between trend and income.

```
. corr cons trend income
(obs=45)

         |     cons    trend   income
---------+---------------------------
    cons |   1.0000
   trend |   0.9783   1.0000
  income |   0.9964   0.9825   1.0000
```

Solutions:

Unfortunately the only sensible thing to do when faced with multicolinearity is either to

Solutions:
Unfortunately the only sensible thing to do when faced with multicolinearity is either to

1) Get more data – (since an increase in N will reduce the standard errors)

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

Solutions:
Unfortunately the only sensible thing to do when faced with multicolinearity is either to

1) Get more data – (since an increase in N will reduce the standard errors)

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

2) Get more (uncorrelated) variables – since this should reduce the residual variance s^2 and offset the multicolinearity effect.

Solutions:
Unfortunately the only sensible thing to do when faced with multicolinearity is either to

1) Get more data – (since an increase in N will reduce the standard errors)

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

2) Get more (uncorrelated) variables – since this should reduce the residual variance s^2 and offset the multicolinearity effect.

3) Change the functional form (try logs rather than polynomials of the same variable)

Solutions:
Unfortunately the only sensible thing to do when faced with multicolinearity is either to

1) Get more data – (since an increase in N will reduce the standard errors)

$$Var(\hat{\beta}_1) = \frac{s^2}{N * Var(X)} * \frac{1}{1 - r^2_{X_1 X_2}}$$

2) Get more (uncorrelated) variables – since this should reduce the residual variance s^2 and offset the multicolinearity effect.

3) Change the functional form (try logs rather than polynomials of the same variable)

If this fails then quite often the only solution is to drop one of the original correlated variables. The issue cannot be addressed given the available data.