ES1004 Econometrics by Example

Lecture 1: Linear Regression Model

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition

April 30th 2016

1. Nature & Structure
2. Estimation
3. Goodness of Fit
4. Testing Hypotheses
5. Illustrative Examples

General Model I

- the general form of the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \tag{1}$$

  - Y dependent, explained, regressand or outcome variable
  - X's independent, explanatory, regressors, predictors or covariates
  - u disturbance, error, stochastic or random term
  - the i subscript denotes the ith observation

General Model II

- for ease of exposition, we write eq. 1 as

$$Y_i = \beta \mathbf{X} + u_i \tag{2}$$

  - $\beta\mathbf{X}$ is a short form of $\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$
  - $\mathbf{X}$ vector of regressors

Population (true) Model I

- eq. 1, or its short form eq. 2, known as the population or true model
  - the term population refers to a well-defined entity (people, firms, cities, states, countries, and so on)
- population (true) model consists of two components
  1. deterministic component $\beta\mathbf{X}$ (the conditional mean of Y ,or $E(Y|X)$
  2. non-systematic, random or stochastic component u_i
- an individual Y_i value is equal to
  1. the mean value of the population of which it is a member
  2. plus or minus a random error term

Population (true) Model II

- for example,
  - if Y represents family expenditure on food, X is family income
  - eq. 2 states that, the food expenditure of an individual family is equal to
    1. the mean food expenditure of all families with the same level of income
    2. plus or minus a random component; varies from individual to individual and depends on several factors

Regression Analysis I

- primary objective is to explain the mean, average, behaviour of Y in relation to the regressors
  - how mean Y responds to changes in the values of the X variables
  - an individual Y value will hover around its mean value
- the causal relationship between Y and the Xs, if any, should be based on the relevant theory

Regression Analysis II

- how many regressors in the model depends on the nature of the problem
- the error term u
  - a catchall for all variables that cannot be introduced in the model
  - the average influence of these variables is assumed to be negligible

Regression Coefficients

- in eq. 1
  - $\beta_0, \beta_1, \ldots, \beta_k$ regression coefficients or parameters
  - β_0 the intercept
  - β_1 to β_k the slope coefficients
  - each slope coefficient measures
    - the (partial) rate of change in the mean value of Y for a unit change in the value of a regressor, holding the values of all other regressors constant (ceteris paribus)

Sample Regression Function

- the sample counterpart is

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} + e_i \tag{3}$$

- or, as written in short form

$$Y_i = \mathbf{bX} + e_i \tag{4}$$

  - where e is a residual

- the deterministic component is written as

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} = \mathbf{bX} \tag{5}$$

Nature of Y

- Y is a random variable and can be measured on four different scales
  1. ratio scale
     - ratio of two variables, distance between two variables, and ordering of variables are meaningful
  2. interval scale
     - distance and order between variables meaningful, but not ratio
  3. ordinal scale
     - ordering of two variables meaningful, but not ratio or distance
  4. nominal scale
     - categorical or dummy variables, qualitative in nature

Time Series Data

- a set of observations that a variable takes at different times, such as
  - daily - stock prices
  - weekly - money supply
  - monthly - the unemployment rate
  - quarterly - gdp
  - annually - government budgets
  - quinquennially or every five years - the census of manufactures
  - decennially or every ten years - the census of population

Cross-Section Data

- data on one or more variables collected at the same point in time
  - opinion polls conducted by various polling organisation
  - temperature at a given time in several places

Pooled Data

- panel, longitudinal or micro panel data
  - combines features of both cross-section and time series data
  - same cross section units are followed over time
  - panel data represents a special type of pooled data
  - pooled data: time series, cross sectional, where the same cross-sectional unit are not necessarily followed over time

OLS Method

- does not minimise the sum of the error term, but minimises error sum of square

$$\sum u_i^2 = \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2 \qquad (6)$$

- to obtain values of the regression coefficients, derivatives are taken with respect to the regression coefficients and set equal to zero

CLRM: Assumptions

A$_1$: model is linear in parameters

A$_2$: regressors are fixed non-stochastic

A$_3$: the expected value of the error term is zero $E(u_i|X) = 0$

A$_4$: homoscedastic or constant variance of errors $var(u_i|X) = \sigma^2$

A$_5$: no autocorrelation, $cov(u_i, u_j) = 0, i \neq j$

A$_6$: no multicollinearity; no perfect linear relationships among the Xs

A$_7$: no specification bias

Gauss-Markov Theorem

- given assumptions $\mathbf{A}_1$ to $\mathbf{A}_7$, OLS gives best linear unbiased estimators BLUE
  1. estimators are linear functions of the dependent variable Y
  2. estimators are unbiased, in repeated samples the estimators approach their true value
  3. in the class of linear estimators, OLS estimators have minimum variance; i.e., they are efficient, or the best estimators

R^2

- an overall measure of goodness of fit of the estimated regression line
- gives the percentage of the total variation in the dependent variable that is explained by the regressors
- it is a value between zero (no fit) and 1 (perfect fit)
  - explained sum of squares $(ESS) = \sum(\hat{Y} - \bar{Y})^2$
  - residual sum of squares $(RSS) = \sum e^2$
  - total sum of squares $(TSS) = \sum(Y - \bar{Y})^2$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{7}$$

t-test

- testing the null hypothesis

$$H_0 : \beta_k = c$$

- against one of three possible alternative hypotheses

$$H_1 : \beta_k > c$$
$$H_1 : \beta_k < c$$
$$H_1 : \beta_k \neq c$$

t-test

- for example, if $c = 0$

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0$$

- calculate test statistic

$$t = \frac{b_k - 0}{se(b_k)} \sim t_{n-K}$$

- reject the null hypothesis if
  - $|t| > t^c_{n-K}$; or
  - $p < \alpha$

F-test

- testing the null hypothesis that all slope coefficients are zero

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

  equivalent to $H_0 : R^2 = 0$
- against the alternative hypothesis that at least one slope coefficient is not zero; $H_1 : R^2 \neq 0$
- calculate the test statistic

$$F = \frac{ESS/df}{RSS/df} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{8}$$

- reject the null hypothesis if
  - $F > F^c$; or
  - $p < \alpha$

F-test

- test $H_0 : \beta_{j1} = \beta_{j2} = \cdots = \beta_{js} = 0$ vs. H_1 : one or another is not zero

$$F = \frac{(R^2_{UR} - R^2_{R})/q}{(1 - R^2_{UR})/n - k} \sim F(q, n - k)$$

  - q the number of zero restrictions
  - k the number of population parameters in the unrestricted model
- reject the null hypothesis if
  - $F > F^c$; or
  - $p < \alpha$

Modelling Wealth

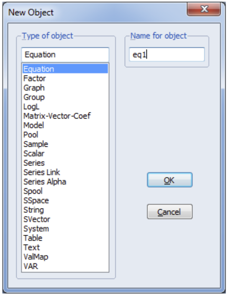
- assume we wish to estimate a regression model that explains people's wealth
    - here wealth is the dependent variable i.e., the variable we want the model to explain
    - factors determine how much wealthy an individual could include income, age, family size, among others
- we collected data on 9275 individuals
    - this data is from wooldridge, introductory econometrics 4^{th} edition

Data

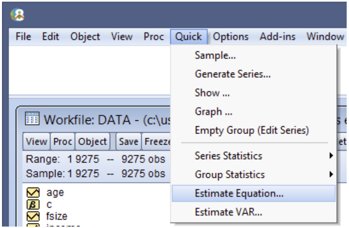
- eviews data file named `wealth-data` on piazza
  - `wealth` – net total financial wealth (in thousands of dollars)
  - `income` – annual income (in thousands of dollars)
  - age – age in years (minimum age in the dataset is 25 years)
  - `fsize` – family size; number of individuals living in the family
- we want to estimate the following model

$$wealth_i = \beta_0 + \beta_1 income_i + \beta_2 age_i + \beta_3 fsize_i + u_i$$

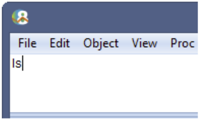
Estimate in Eviews



1. From the Main menu, select **Object** → **New Object** → **Equation.**

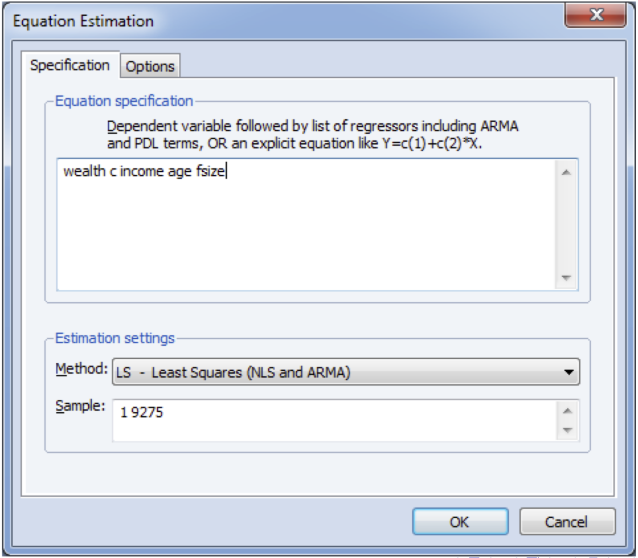


2. From the Main menu, select **Quick** → **Estimate Equation.**



3. On the command window type: `ls`

Estimate in Eviews



Estimation output

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Reporting Results

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

Interpretation

Equation: UNTITLED   Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13   Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Interpretation

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- intercept $\hat{\beta}_0$
  - wealth equals - 52.674 (thousands of dollars), on average, when income, age and family size equal zero
  - does it make sense?

Interpretation

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Interpretation

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

slope coefficients

- $\hat{\beta}_1$ `income`
  - when income increases by 1 unit (a thousand dollar), wealth increases, on average, by 0.974 unit ($974); assuming that age and family size did not change

Interpretation

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Interpretation

$$\widehat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

slope coefficients

- $\hat{\beta}_2$ age
  - when age increases by one unit (a year), wealth increases on average, increases by 1.013 units ($1013); assuming income and family size are held constant

Interpretation

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Interpretation

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

slope coefficients

- $\hat{\beta}_3$ `fsize`
  - when number of individuals living in the family increases by one unit (person), wealth on average decreases by 2.806 units ($2806); holding other factors constant

Goodness of Fit

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Goodness of Fit

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- $R^2 = 0.1734$
  - the model explains 17.3% of the variation in the net total financial wealth
  - the ‘model’ here means income, age and family size together
- adjusted $R^2 = 0.1732$
  - interpreted in the same way
  - adjust for degrees of freedom

Testing Hypotheses: Individual Parameters

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- assume we want to test whether income has statistical significant impact on an individual's net wealth
  - $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$
  - employ a t-test

$$t = \frac{b_1 - \beta_1}{se(b_1)} \sim t_{n-k}$$

Testing Hypotheses: Individual Parameters

$$\widehat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

$$t = \frac{0.973735 - 0}{0.025377} = 38.3707688 \quad \sim t_{[9275-4=9271]}$$

- 2-t statistic rule: reject the null hypothesis
- income has a statistical significant impact on an individual's net wealth

Testing Hypotheses: t-ratio

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Reporting Results: t-ratios

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Reporting Results: t-ratios

$$\widehat{wealth}_i = \underset{[-18.617]}{-52.674} + \underset{[38.370]}{0.974}\ income_i + \underset{[17.164]}{1.013}\ age_i - \underset{[-7.039]}{2.806}\ fsize_i$$

t statistic in brackets

Reporting Results: t-ratios

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Reporting Results: p-values

$$\hat{wealth}_i = \underset{(0.000)}{-52.674} + \underset{(0.000)}{0.974}\ income_i + \underset{(0.000)}{1.013}\ age_i - \underset{(0.000)}{2.806}\ fsize_i$$

p-value in brackets

Testing Hypotheses: Joint Hypotheses

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- what if we wish to test whether income, age and family size collectively have no significant impact on an individual's net wealth
  - $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
  - i.e., testing the significance of the model $H_0 : R^2 = 0$
  - employ an F-test

Testing Hypotheses: Joint Hypotheses

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- what if we wish to test whether income, age and family size collectively have no significant impact on an individual's net wealth
  - $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

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

Testing Hypotheses: Joint Hypotheses

$$\widehat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- what if we wish to test whether income, age and family size collectively have no significant impact on an individual's net wealth
  - $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$$F = \frac{0.173474/(4-1)}{(1-0.173474)/(9275-4)} = 648.6115 \sim F_{(4-1,9275-4)}$$

Testing Hypotheses: Joint Hypotheses

F - Distribution ($\alpha = 0.05$ in the Right Tail)

df_2 \ df_1 (Denominator Degrees of Freedom \ Numerator Degrees of Freedom)	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

Testing Hypotheses: Joint Hypotheses

$$\hat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- what if we wish to test whether income, age and family size collectively have no significant impact on an individual's net wealth
  - $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
- since that $F > F^c$, we reject the null hypothesis; the model is significant

Testing Hypotheses: Joint Hypotheses

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Testing Hypotheses: Joint Hypotheses

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13 Time: 09:07
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

Testing Hypotheses: Exclusion Restrictions

$$\widehat{wealth}_i = \underset{(2.829)}{-52.674} + \underset{(0.025)}{0.974}\ income_i + \underset{(-.059)}{1.013}\ age_i - \underset{(0.399)}{2.806}\ fsize_i$$

standard errors in brackets

- what if we wish to test whether it is important to include age and family size in this model
  - $H_0 : \beta_2 = \beta_3 = 0$
  - i.e., testing whether they can be excluded from this model
  - employ an F-test

Testing Hypotheses: Exclusion Restrictions

- unrestricted model

$$wealth_i = \beta_0 + \beta_1 income_i + \beta_2 age_i + \beta_3 fsize_i + u_i$$

$$R^2_{UR} = 0.173474$$

$$H_0 : \beta_2 = \beta_3 = 0$$

- restricted model

$$wealth_i = \beta_0 + \beta_1 income_i + u_i$$

$$R^2_R = 0.141817$$

Testing Hypotheses: Exclusion Restrictions

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 02/29/16 Time: 14:51
Sample: 1 9275
Included observations: 9275

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-20.17948	1.176434	-17.15309	0.0000
INCOME	0.999911	0.025543	39.14569	0.0000

R-squared	0.141817	Mean dependent var	19.07168
Adjusted R-squared	0.141724	S.D. dependent var	63.96384
S.E. of regression	59.25813	Akaike info criterion	11.00190
Sum squared resid	32562380	Schwarz criterion	11.00344
Log likelihood	-51019.30	Hannan-Quinn criter.	11.00242
F-statistic	1532.385	Durbin-Watson stat	1.939237
Prob(F-statistic)	0.000000		

Testing Hypotheses: Exclusion Restrictions

$$H_0 : \beta_2 = \beta_3 = 0$$

$$F = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/n - k} \qquad \sim F_{(q,n-k)}$$

$$F = \frac{(0.173474 - 0.141817)/2}{(1 - 0.173474)/9275 - 4} = 177.5496 \qquad \sim F_{(2,9271)}$$

- since that $F > F^c$, we reject the null hypothesis
- age and `fsize` should be included in the model (can not be excluded)

Testing Hypotheses: Exclusion Restrictions

Equation: UNTITLED   Workfile: DATA::Estimation\

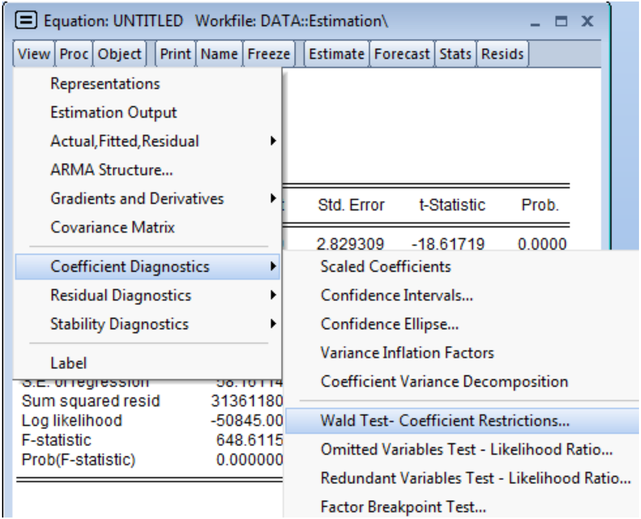
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WEALTH
Method: Least Squares
Date: 03/04/13   Time: 09:07
Sample: 1 9275
Included observations: 9275

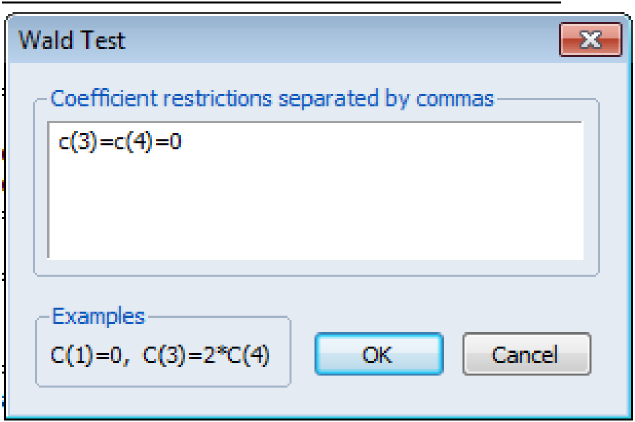
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-52.67379	2.829309	-18.61719	0.0000
INCOME	0.973735	0.025377	38.37042	0.0000
AGE	1.013050	0.059022	17.16389	0.0000
FSIZE	-2.805637	0.398602	-7.038697	0.0000

R-squared	0.173474	Mean dependent var	19.07168
Adjusted R-squared	0.173207	S.D. dependent var	63.96384
S.E. of regression	58.16114	Akaike info criterion	10.96474
Sum squared resid	31361180	Schwarz criterion	10.96782
Log likelihood	-50845.00	Hannan-Quinn criter.	10.96579
F-statistic	648.6115	Durbin-Watson stat	1.929874
Prob(F-statistic)	0.000000		

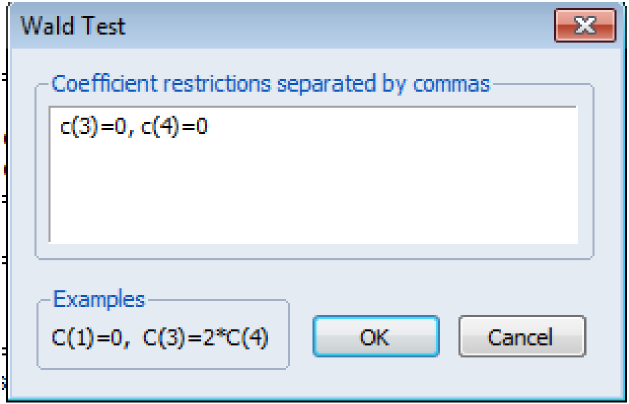
Testing Hypotheses: Exclusion Restrictions



Testing Hypotheses: Exclusion Restrictions



Testing Hypotheses: Exclusion Restrictions



Testing Hypotheses: Exclusion Restrictions

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	177.5496	(2, 9271)	0.0000
Chi-square	355.0992	2	0.0000

Null Hypothesis: C(3)=C(4)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	1.013050	0.059022
C(4)	-2.805637	0.398602

Restrictions are linear in coefficients.

Testing Hypotheses: Exclusion Restrictions

Equation: UNTITLED Workfile: DATA::Estimation\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	177.5496	(2, 9271)	0.0000
Chi-square	355.0992	2	0.0000

Null Hypothesis: C(3)=0, C(4)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	1.013050	0.059022
C(4)	-2.805637	0.398602

Restrictions are linear in coefficients.

مجتمع الاقتصاد THE ECONOMIC SOCIETY