ES1004 Econometrics by Example

Lecture 3: Qualitative Explanatory Variables

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition

14th May 2016

Basic Idea I

- sometimes cannot obtain set of numerical values for all variables to use in a model
- because some variables cannot be quantified easily

- examples
  - gender may play a role in determining salary levels
  - different ethnic groups may follow different consumption patterns
  - educational levels can affect earnings from employment

Basic Idea II

- qualitative variables as regressors
  - to include in a regression we define dummy variables
  - nominal scale variables which have no particular numerical values
  - usually in cross-sectional models, but can appear in time series as well

- more examples [in times series]
  - changes in political regime may affect production
  - war can impact on economic activities
  - certain days in week or certain months in year can have different effects on the fluctuation of stock prices
  - seasonal effects often observed in demand of various products

Basic Idea III

- note that dummy variables are also called
  - indicator variables
  - categorical variables, and
  - qualitative variables

Including Dummy Variables I

- consider following cross-sectional model

$$wage_i = \beta_1 + \beta_2 exper_i + u_i$$

- this model assumes that the constant β_1 is the same for all the observations in our dataset
- what if we have two different subgroups
  - male and female, for example

Including Dummy Variables II

- we convert such qualitative information into a quantitative variable by creating a dummy variable

$$D = \begin{cases} 1 & \text{if female} \\ 0 & \text{if male} \end{cases}$$

- note that
  - i the choice of which of the two different outcomes is to be assigned the value of 1 does not alter the results
  - ii the 0 classification is often referred to as the benchmark, or control category

Including Dummy Variables III

$$wage_i = \beta_1 + \beta_2 exper_i + \beta_3 D_i + u_i$$

- now we have two cases
  1. when $D = 0$ (male)
  $$Y_i = \beta_1 + \beta_2 X_i + u_i$$
  2. when $D = 1$ (female)
  $$Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + u_i$$
- two groups [male & female] but included only one dummy D_i

Dummy Variable Trap

- if an intercept is included and we have a qualitative variable with m categories, then introduce only $(m-1)$ dummy variables
  - consider a self-reported health as a choice among excellent, good, and poor
  - we can have at most two dummy variables to represent three categories
- not following this rule $\rightarrow$ dummy variable trap $\rightarrow$ perfect collinearity

Reference Category

- the subgroup that gets value of 0 is called the
  - reference category,
  - benchmark, or
  - comparison category

- all comparisons are made in relation to the reference category
  - if there are several dummy variables, you must keep track of the reference category

Modelling Wages: Data

- we want to study what factors determine hourly wage (in dollars)
- table 1.1 data of 1289 individuals interviewed in March 1995

  - `wage` ☞ hourly wage in dollars [dependent variable]
  - `female` ☞ gender, coded 1 for female, 0 for male
  - `nonwhite` ☞ race, coded 1 for nonwhite, 0 for white workers
  - `union` ☞ union status, coded 1 if in a union job, 0 otherwise
  - `education` ☞ education in years
  - `exper` ☞ potential work experience in years

Modelling Wages: Estimation

Dependent Variable: WAGE
Method: Least Squares
Date: 05/14/16 Time: 09:39
Sample: 1 1289
Included observations: 1289

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.183338	1.015788	-7.071691	0.0000
FEMALE	-3.074875	0.364616	-8.433184	0.0000
NONWHITE	-1.565313	0.509188	-3.074139	0.0022
UNION	1.095976	0.506078	2.165626	0.0305
EDUCATION	1.370301	0.065904	20.79231	0.0000
EXPER	0.166607	0.016048	10.38205	0.0000

Refining the Wage Function I

- we found that the average salary of a
  - female worker is lower than that of her male counterpart
  - nonwhite worker is lower than that of his white counterpart

- what about a female nonwhite?
- we need to include an interactive dummy

Refining the Wage Function I

Dependent Variable: WAGE
Method: Least Squares
Date: 05/14/16 Time: 10:19
Sample: 1 1289
Included observations: 1289

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.088725	1.019482	-6.953264	0.0000
FEMALE	-3.240148	0.395328	-8.196106	0.0000
NONWHITE	-2.158525	0.748426	-2.884087	0.0040
UNION	1.115044	0.506352	2.202113	0.0278
EDUCATION	1.370113	0.065900	20.79076	0.0000
EXPER	0.165856	0.016061	10.32631	0.0000
FEMALE*NONWHITE	1.095371	1.012897	1.081424	0.2797

Refining the Wage Function II

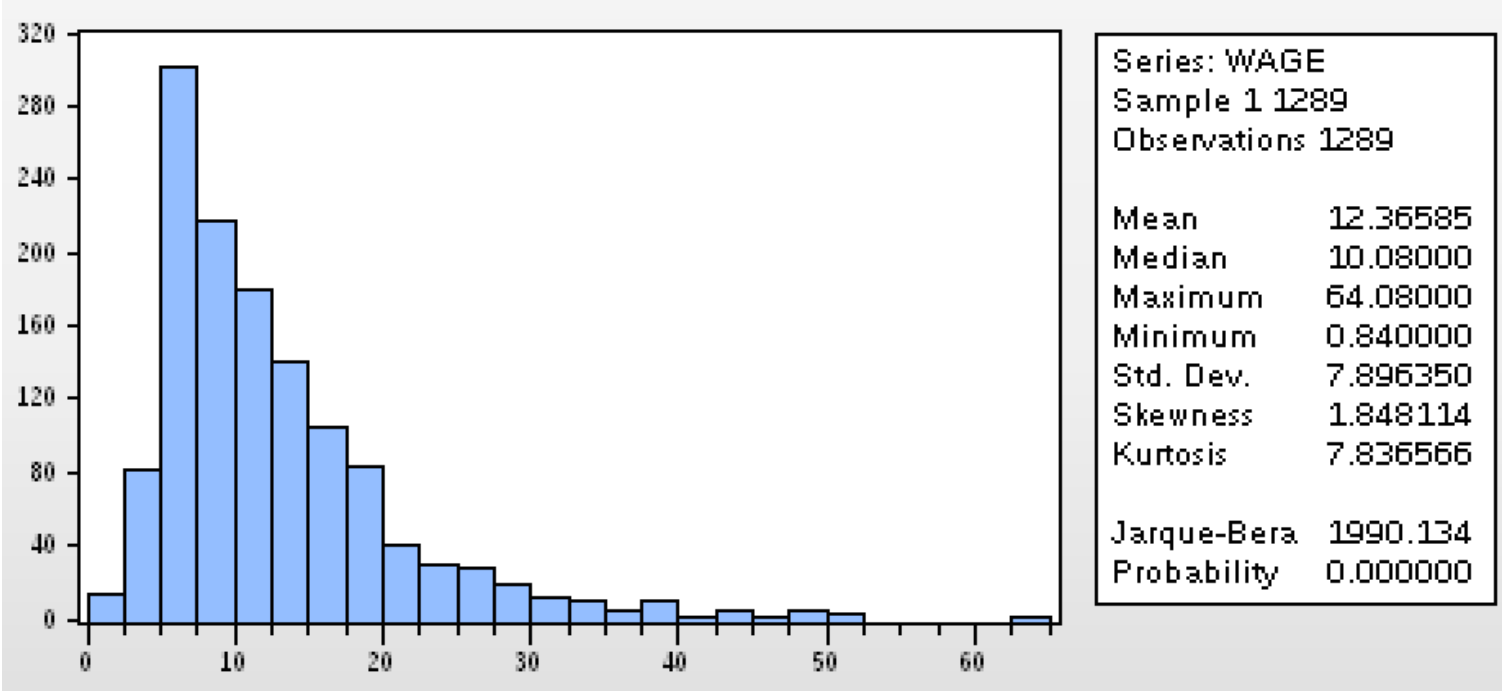
- we implicitly assumed that slope coefficients of quantitative regressors remain the same between
  - male and female
  - white and nonwhite
- however, we do not need to - we can include differential slope dummies

Refining the Wage Function II

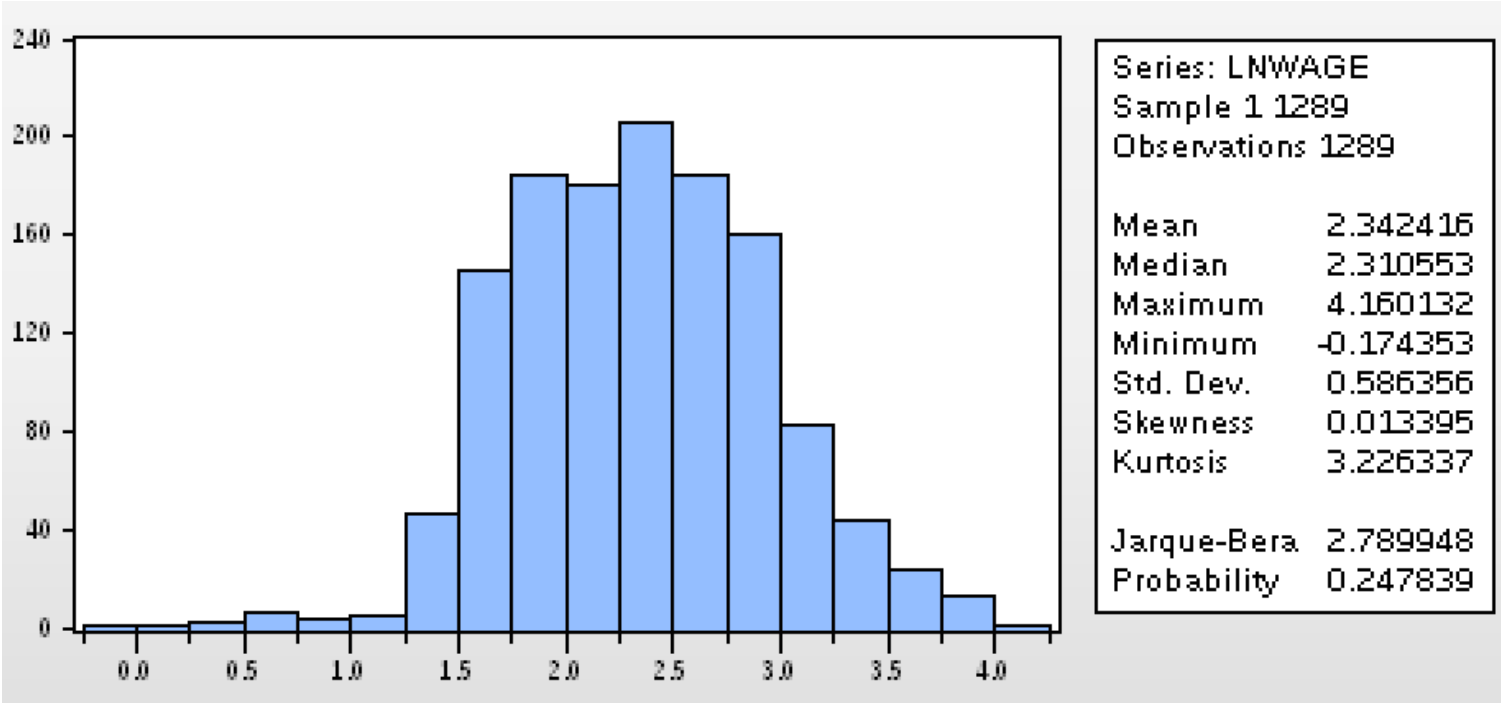
Dependent Variable: WAGE
Method: Least Squares
Date: 05/14/16 Time: 10:23
Sample: 1 1289
Included observations: 1289

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11.09129	1.421846	-7.800623	0.0000
FEMALE	3.174158	1.966465	1.614144	0.1067
NONWHITE	2.909129	2.780066	1.046424	0.2956
UNION	4.454212	2.973494	1.497972	0.1344
EDUCATION	1.587125	0.093819	16.91682	0.0000
EXPER	0.220912	0.025107	8.798919	0.0000
FEMALE*EDUCATION	-0.336888	0.131993	-2.552314	0.0108
FEMALE*EXPER	-0.096125	0.031813	-3.021530	0.0026
NONWHITE*EDUCATI...	-0.321855	0.195348	-1.647595	0.0997
NONWHITE*EXPER	-0.022041	0.044376	-0.496700	0.6195
UNION*EDUCATION	-0.198323	0.191373	-1.036318	0.3003
UNION*EXPER	-0.033454	0.046054	-0.726410	0.4677

Functional Form



Functional Form



Functional Form

Dependent Variable: LNWAGE
Method: Least Squares
Date: 05/14/16 Time: 10:34
Sample: 1 1289
Included observations: 1289

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.905504	0.074175	12.20768	0.0000
FEMALE	-0.249154	0.026625	-9.357891	0.0000
NONWHITE	-0.133535	0.037182	-3.591399	0.0003
UNION	0.180204	0.036955	4.876316	0.0000
EDUCATION	0.099870	0.004812	20.75244	0.0000
EXPER	0.012760	0.001172	10.88907	0.0000