Erasmus
School of
Economics

MOOC Econometrics

Lecture 2.5 on Multiple Regression: Application

Christiaan Heij

Erasmus University Rotterdam

Wage equation

- Wage data of Lecture 2.1 with model of Lecture 2.2.
- Model: $\log(\text{Wage})_i = \beta_1 + \beta_2 \text{Female}_i$
$$+\beta_3 \text{Age}_i + \beta_4 \text{Educ}_i + \beta_5 \text{Parttime}_i + \varepsilon_i$$
- OLS gives: $\log(\text{Wage})_i = 3.05 - 0.04\text{Female}_i$
$$+0.03\text{Age}_i + 0.23\text{Educ}_i - 0.37\text{Parttime}_i + e_i$$
- $R^2 = 0.704$ and $s = 0.245$.
- Data are random sample from population of employees.

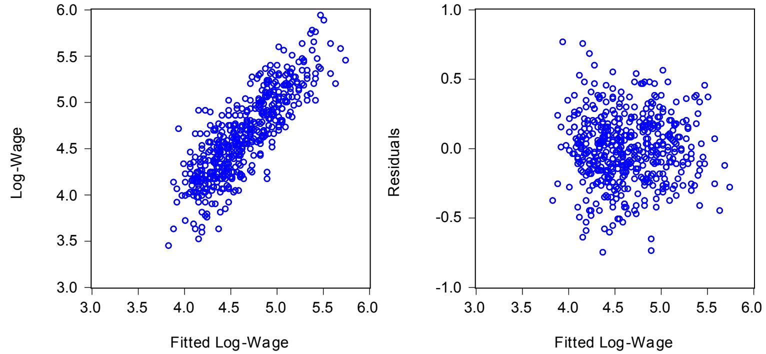
  OLS results depend on these data, hence also random.

Regression outcomes

Dependent variable: log(Wage)
Sample size: 500

	Coefficient b_j	Standard error $SE(b_j)$	t-Statistic t_j	p-value $H_0 : \beta_j = 0$
Constant	3.053	0.055	55.168	0.000
Female	-0.041	0.025	-1.663	0.097
Age	0.031	0.001	24.041	0.000
Educ	0.233	0.011	21.874	0.000
Parttime	-0.365	0.032	-11.576	0.000
R-squared	0.704			
SE of regression	0.245			

Two scatter diagrams



- Left diagram: Actual log-wage against fitted log-wage.
- Right diagram: Residuals against fitted log-wage.

Regression outcomes

- Age, Education, and Parttime are significant (p-values 0.000).

  Female is not significant at 5% level (p-value 0.097).

- Interpretation in terms of average wage effects:

  Extra year of age: $e^{0.031} - 1 = 3\%$

  Extra level of education: $e^{0.233} - 1 = 26\%$

  Part-time job: $e^{-0.365} - 1 = -31\%$

- After controlling for age, education, and part-time job effects, the (partial) gender effect of -4% for females is not significant.

- Lecture 2.1: Significant gender effect of -25% for females: total effect, including education and part-time jobs.

Model with absolute (instead of relative) effects

- If explained variable is log-wage, parameters

$$\beta_j = \partial \log(\text{Wage})/\partial x_j = (\partial \text{Wage}/\partial x_j)/\text{Wage}$$

  measure relative wage effects of each factor.

- If explained variable is wage (instead of log-wage), parameters $\beta_j = \partial \text{Wage}/\partial x_j$ measure wage level effects.

- OLS in this model gives:

$$\text{Wage}_i = -77.87 - 2.12\text{Female}_i$$
$$+3.62\text{Age}_i + 29.47\text{Educ}_i - 43.10\text{Parttime}_i + e_i.$$

- $R^2 = 0.681$ and $s = 31.276$.

Wage or log-wage?

- Age, Education, and Parttime are significant (p-values 0.000).

  Female is not significant at 5% level (p-value 0.501).

Test

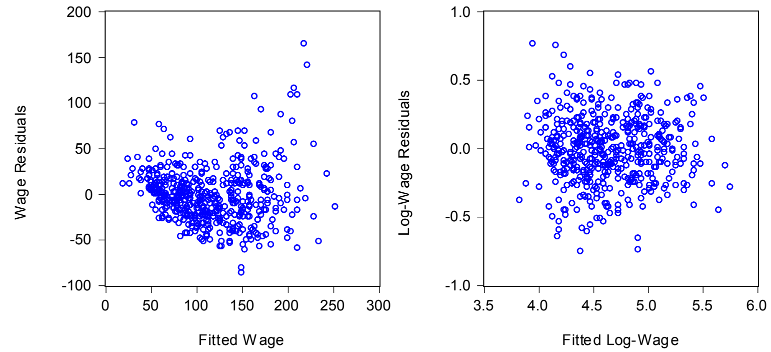
Why can we not choose between the two models (with log-wage and wage) on the basis of R^2 and s?

- Answer: R^2 and s are based on sum of squares of y and $e = y - Xb$, and y differs in the two models.

- Graphical check of regression assumptions:

  scatter diagram of residuals against fitted values.

Scatter diagrams of residuals against fitted values



- Left for wage: nonlinear and heteroskedastic

- Right for log-wage: no indication violation regression assumptions

Testing for constant education effects

- Allow that education effect varies per education level:

$$\log(\text{Wage})_i = \beta_1 + \beta_2 \text{Female}_i + \beta_3 \text{Age}_i$$
$$+\beta_4 \text{DE2}_i + \beta_5 \text{DE3}_i + \beta_6 \text{DE4}_i + \beta_7 \text{Parttime}_i + \varepsilon_i$$

- $\text{DE2}_i = 1$ if employee i has education level 2
  $\text{DE2}_i = 0$ if employee i has education level 1, 3, or 4

  (similar definitions for DE3 and DE4)

- Effect of education is constant if $\beta_5 = 2\beta_4$ and $\beta_6 = 3\beta_4$.

- Test $H_0 : \beta_5 = 2\beta_4$ and $\beta_6 = 3\beta_4$ against H_1: H_0 not true.

Regression outcomes

Dependent variable: log(Wage)
Sample size: 500

	Coefficient b_j	Standard error $SE(b_j)$	t-Statistic t_j	p-value $H_0 : \beta_j = 0$
Constant	3.318	0.051	64.554	0.000
Female	-0.031	0.024	-1.267	$\boxed{0.206}$
Age	0.030	0.001	24.269	0.000
DE2	0.171	0.027	6.308	0.000
DE3	0.380	0.029	12.996	0.000
DE4	0.767	0.035	21.610	0.000
Parttime	-0.366	0.031	-11.813	0.000
R-squared	0.716			
SE of regression	0.241			

Outcomes

- OLS in unrestricted model (under H_1) gives:

$$\log(\text{Wage})_i = 3.32 - 0.03\text{Female}_i + 0.03\text{Age}_i$$
$$+0.17\text{DE2}_i + 0.38\text{DE3}_i + 0.77\text{DE4}_i - 0.37\text{Parttime}_i + e_i.$$

- $R^2 = 0.716$ and $s = 0.241$.

- All factors are significant, except for 'Female' (p-value 0.206).

- Test for constant education effects:

  Test $H_0 : \beta_5 = 2\beta_4, \beta_6 = 3\beta_4$ against H_1: H_0 not true.

Test

Compute the F-test, using $R_1^2 = 0.716$ and $R_0^2 = 0.704$.

Computation of F-test

- $R_1^2 = 0.716$ and $R_0^2 = 0.704$

- $g = 2$, $n = 500$, $k = 7$ (under H_1), $n - k = 500 - 7 = 493$

  $F = \frac{(R_1^2 - R_0^2)/g}{(1-R_1^2)/(n-k)} = \frac{(0.716-0.704)/2}{(1-0.716)/493} = 10.4$

- 5% critical value of $F(2, 493)$ is 3.0.

  As $F = 10.4 > 3.0$, H_0 is rejected (at 5% level).

- Conclusion: Wage effect of one extra level of education differs significantly across education levels.

Wage effect of extra education

- Coefficients of education dummies:

  level 2: 0.171

  level 3: 0.380

  level 4: 0.767

- Wage increase for higher education level:

  $1 \to 2$: $e^{0.171} - 1 = 0.19 = 19\%$

  $2 \to 3$: $e^{(0.380-0.171)} - 1 = e^{0.209} - 1 = 23\%$

  $3 \to 4$: $e^{(0.767-0.380)} - 1 = e^{0.387} - 1 = 47\%$

- Effect much larger for highest education level.

TRAINING EXERCISE 2.5

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).