

Multiple Regression Analysis : Further Issues

1 Data scaling on OLS statistics

When we change the unit of measurement of a variable, the value of estimators would change accordingly. For example

$$\widehat{bweght} = \widehat{\beta}_0 + \widehat{\beta}_1 cigs + \widehat{\beta}_2 fa\ min\ c,$$

where

bwght = child birth weight, in grams.

cigs = number of cigarettes smoked by the mother while pregnant, per day.

fa min c = annual family income, in thousands of dollars.

2 More on functional forms

- Logarithmic Functional Form

$$\log(y) = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + u$$

- Models with Quadratics

Example : Effects of Pollution on Housing Prices

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \log(\text{dist}) + \beta_3 \text{rooms} + \beta_4 \text{room}^2 + \beta_5 \text{stratio} + u$$

where

- price* = housing price
- nox* = level of pollution
- dist* = distance from downtown
- rooms* = number of rooms
- stratio* = average student per teacher ratio

The estimation result is given by

regress lprice lnox dist rooms rooms_sq stratio

Source	SS	df	MS			
Model	51.4933152	5	10.298663	Number of obs =	506	
Residual	33.0889098	500	.06617782	F(5, 500) =	155.62	
Total	84.582225	505	.167489554	Prob > F =	0.0000	
				R-squared =	0.6088	
				Adj R-squared =	0.6049	
				Root MSE =	.25725	

lprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnox	-.9767545	.0995938	-9.81	0.000	-1.172429	-.7810806
dist	-.0321972	.0094013	-3.42	0.001	-.050668	-.0137264
rooms	-.5528032	.1612965	-3.43	0.001	-.8697056	-.2359007
rooms_sq	.0624697	.0124867	5.00	0.000	.0379368	.0870025
stratio	-.0486667	.0058131	-8.37	0.000	-.0600879	-.0372455
_cons	13.59154	.5650901	24.05	0.000	12.4813	14.70178

Consider the effect of "room"

What would be the % change in price when the number of room increases from 5 to 6?

3 Models with Interaction Terms

Consider

$$price = \beta_0 + \beta_1 sqft + \beta_2 bdrms + \beta_3 sqft \times bdrms + \beta_4 bthrms + u$$

where

price = housing price

sqft = house size (square feet)

bdrms = number of bedrooms

bthrms = number of bathrooms

4 More on the Goodness-of-Fit and Selection of Regressors

- Adding more regressors ALWAYS improve fit

Using adjusted R-squared to choose between non-nested models (one model is not a subset of another).

Consider Model 1

$$\begin{aligned} \widehat{salary} &= 830.63 + 0.0163sales + 19.63roe \\ &\quad (223.90) \quad (0.0089) \quad (11.08) \\ n &= 209, \quad R^2 = 0.029, \quad \bar{R}^2 = 0.020 \end{aligned}$$

Consider Model 2

$$\begin{aligned} \log(\widehat{salary}) &= 4.36 + 0.2751 \log(sales) + 0.0179roe \\ &\quad (0.29) \quad (0.033) \quad (0.004) \\ n &= 209, \quad R^2 = 0.282, \quad \bar{R}^2 = 0.275 \end{aligned}$$

