Multiple Regression Analysis : Further Issues

1 Data scaling on OLS statistics

When we change the unit of measurment of a variable, the value of estimators would chance 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 Pullution on Housing Prices

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

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
				Prob > F =	0.0000
				R-squared =	0.6088
				Adj R-squared =	0.6049
Total	84.582225	505	.167489554	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 wold be the % change in price when the number of room increases from 5 to 6?

3 Models with Interaction Terms

Consider

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

where
$price$ = housing price
$sqrft$ = 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

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

Consider Model 2

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