ECON 3740: INTRODUCTION TO ECONOMETRICS

INSTRUCTOR: CHAOYI CHEN
Department of Economics and Finance, University of Guelph

Lecture 16

Lecture outline

Last lecture, we studied the model with quadratic term and the interaction term. Today, we will

- Learn adjusted R^2
- Study how to choose between nonnested models.
- Study when we should add regressor into the model
- Predict y when $log(y)$ is the dependent variable

MLR, Further Issue: Adjusted R-Squared

- General remarks on R-squared:
  - A high R-squared does not imply that there is a causal interpretation.
  - A low R-squared does preclude precise estimation of partial effects.
- Recall that

$$R^2 = 1 - \frac{SSR/n}{SST/n} = 1 - \frac{\tilde{\sigma}_\mu^2}{\tilde{\sigma}_y^2},$$

  so R^2 is estimating the population R-squared

$$\rho^2 = 1 - \frac{\sigma_\mu^2}{\sigma_y^2},$$

  the proportion of the variation in y in the population explained by the independent variables.
- Adjusted R-Squared:

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\widehat{\sigma}_\mu^2}{\widehat{\sigma}_y^2}$$

  is sometimes also called R-bar squared, where $\widehat{\sigma}_\mu^2$ and $\widehat{\sigma}_y^2$ are unbiased estimators of σ_μ^2 and σ_y^2 due to the correction of dfs.

MLR, Further Issue: Adjusted R-Squared Continue

- $\bar{R}^2$ takes into account degrees of freedom of the numerator and denominator, so is generally a better measure of goodness-of-fit.
- $\bar{R}^2$ imposes a penalty for adding new regressors: $k \uparrow \Longrightarrow \bar{R}^2 \downarrow$
- $\bar{R}^2$ increases if and only if the t statistic of a newly added regressor is greater than one in absolute value. For example, compare with $y = \beta_0 + \beta_1 x_1 + \mu$, the regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$ has a larger $\bar{R}^2$ if and only if

$$\left| t_{\widehat{\beta}_2} \right| > 1$$

- Relationship between R^2 and $\bar{R}^2$,

$$1 - R^2 = \frac{SSR}{SST} = \frac{n-k-1}{n-1} \frac{SSR/(n-k-1)}{SST/(n-1)} = \frac{n-k-1}{n-1}(1 - \bar{R}^2)$$

we have, if $k \neq 0$ and $R^2 < 1$,

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1} < R^2$$

- Note that $\bar{R}^2$ even gets negative if $R^2 < \frac{k}{n-1}$

MLR, Further Issue: Choice between Nonnested Models

- Models are nonnested if neither model is a special case of the other.
- For example, to incorporate diminishing return of *sales* to *R&D*, we consider two models,

$$rdintens = \beta_0 + \beta_1 log(sales) + \mu$$
$$rdintens = \beta_0 + \beta_1 sales + \beta_2 sales^2 + \mu$$

  where *rdintens* is *R&D* intensity.
- Now, suppose after estimate both models, we have $R^2 = 0.061$ and $\bar{R}^2 = 0.03$ in model 1, $R^2 = 0.148$ and $\bar{R}^2 = 0.09$ in model 2.
- A comparison between the R-squared of both models would be unfair to the first model because the first model contains fewer parameters.
- However, even after adjusting for the difference in degrees of freedom, the quadratic model is preferred. Therefore, we may consider choosing model 2.

MLR, Further Issue: Comparing Models with Different Dependent Variables

- R-squared or adjusted R-squared must not be used to compare models which differ in their definition of the dependent variable.
- An example: One would like to investigate the effect of firm performance on CEO salary. Therefore, after estimation, we have TWO fitted regression line based on two Dependent Variables

$$\begin{array}{llll} \widehat{salary} = & 830.63 & +0.0163\text{sales} & +19.63\text{roe} \\ & (223.9) & (0.0089) & (11.08) \end{array}$$

where $n = 209$, $R^2 = 0.029$, $\bar{R}^2 = 0.02$, $SST = 391,732,982$ and

$$\begin{array}{llll} \widehat{log(salary)} = & 4.36 & +0.275\log(\text{sales}) & +0.0179\text{roe} \\ & (0.29) & (0.033) & (0.004) \end{array}$$

where $n = 209$, $R^2 = 0.282$, $\bar{R}^2 = 0.275$, $SST = 66.72$

- There is much less variation in $log(salary)$ that needs to be explained than in salary, so it is not fair to compare R^2 and $\bar{R}^2$ of the two models. (we will discuss how to compare the fitting of these two models later).

MLR, Further Issue: Controlling for Too Many Factors in Regression Analysis

- In some cases, certain variables should not be held fixed:
  - In a regression of traffic fatalities on state beer taxes (and other factors) one should not directly control for beer consumption.
  - why? Beer taxes influence traffic fatalities only through beer consumption
  - In a regression of family health expenditures on pesticide usage among farmers one should not control for doctor visits.
  - why? Health expenditures include doctor visits, an we would like to pick up all effects of pesticide use on health expenditure.
- Different regressions may serve different purposes:
  - In a regression of house prices on house characteristics, one would include price assessments and also housing attributes if the purpose of the regression is to study the validity of assessments; one should not include price assessments if the purpose of the regression is to estimate a hedonic price model, which measures the marginal values of various housing attributes.