ECON 3740: INTRODUCTION TO ECONOMETRICS

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

Lecture 15

Lecture outline

Last lecture, we studied the the hypothesis test with multiple restrictions. Today, we will

- Test general linear restrictions
- Summary: MLR, Inference
- MLR, Further Issues
    - Models with Quadratics
    - Models with Interation Terms
    - Models with Reparametrization of Interaction Effects

MLR: Testing Multiple Linear Restrictions: Testing General Linear Restrictions

- Suppose you and your group member would like to test whether house price assessments are rational, where the population regression model is

$$\begin{aligned} log(price) = \beta_0 + \beta_1 log(assess) + \beta_2 log(lotsize) \\ +\beta_3 log(sqrft) + \beta_4 bdrms + \mu \end{aligned}$$

where

$$\begin{aligned} price &= \text{house price} \\ assess &= \text{the assessed housing value (before the house was sold)} \\ lotsize &= \text{size of the lot, in feet} \\ sqrft &= \text{square footage} \\ bdrms &= \text{number of bedrooms} \end{aligned}$$

- Now, suppose we focus on the following hypothesis test

$$H_0 : \beta_1 = 1, \beta_2 = \beta_3 = \beta_4 = 0$$

MLR: Testing Multiple Linear Restrictions: Testing General Linear Restrictions

- Under the null, $\beta_1 = 1$, which means that if house price assessments are rational, a 1% change in the assessment should be associated with a 1% change in price. (log-log model). Also, $\beta_2 = \beta_3 = \beta_4 = 0$, which means that in addition, other known factors should not influence the price once the assessed value has been controlled for.
- Since it is a test involving multiple linear restrictions, we consider to use F test.
- The restricted model is

$$y = \beta_0 + x_1 + \mu$$

- However, if you just regress y on x_1, we cannot ensure the restriction $\beta_1 = 1$ hold. Hence, we consider to change the resticted model as $y - x_1 = \beta_0 + \mu$, which means The restricted model is actually a regression of $y - x1$ on a constant, and the resulting $\widehat{\beta}_0$ is the sample mean of $y - x_1$.

MLR: Testing Multiple Linear Restrictions: Testing General Linear Restrictions Continue

- Suppose that after estimation with both unrestricted model restricted model, you have the following results

$$SSR_r = 1.88, \; SSR_{ur} = 1.822, \; n = 88$$

- Hence, the F statistic is

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{(1.88 - 1.822)/4}{1.822/(88-4-1)} \approx 0.661$$

- Checking the F table, you find the critical value with 5% significance level of a $F_{4,83}$ distribution is 2.5, which is greater than 0.661. Therefore, we cannot reject the null

MLR: Inference Summary

- In this topic, we have introduced both single restriction test and multiple restrictions test about the MLR.
- First, we showed that, with normality assumption, the OLS estimator $\widehat{\beta}_j$ follows a normal distribution with mean β_j and varriance $Var(\widehat{\beta}_j)$. Moreover, if we standardize the $\widehat{\beta}_j$, the standardized OLS estimator follows a standar normal distribution.
- However, the $sd(\widehat{\beta}_j)$ depends upon an unkonwn parameter σ^2. By replacing $sd(\widehat{\beta}_j)$ with $se(\widehat{\beta}_j)$, we can show that the normal distribution is replaced by a t-distribution.
- Therefore, we can construct t statistic to conduct the hypothesis test. We have shown the hypothesis tests about one parameter with both one sided test and two sided test. We also have shown how to compute P value for a t test and how to construct a confidence interval.

MLR: Inference Summary Continue

- Then, we moved to focus on a hypothesis test with many restrictions.
- We introduced the concepts about the restriction model and unrestriction model.
- To evaluate a multiple restriction hypothesis test, we can use a F test. The statistic can be computed by using the SSR under both restriction model and unrestriction model. Additionally, we showed that we can use a R squared form if a F statistic.
- Lastly, we introduced a generalized linear restriction hypothesis test.

MLR: Models with Quadratics

- Consider we have a fitted regression line for the wage equation as following:

$$\widehat{wage} = \underset{(0.35)}{3.73} + \underset{(0.041)}{0.298 exper} - \underset{(0.0009)}{0.0061 exper^2}$$

  where $n = 526$, and $R^2 = 0.093$
- The predicted wage is a concave function of exper.
- The marginal effect of exper on *wage* is

$$\frac{\partial wage}{\partial exper} = \widehat{\beta}_1 + 2\widehat{\beta}_2 exper = 0.298 - 2 * 0.0061 exper$$

- The first year of experience increases the wage by some \$0.3, the second year by $0.298 - 2 * (0.0061) * 1 = \$0.29 < \$0.3$

MLR: Models with Quadratics Continue

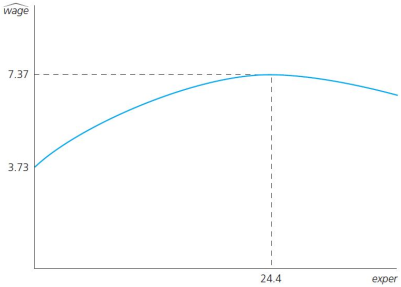

Figure: Wage Maximum w.r.t Work Experience

- Does this mean the return to experience becomes negative after 24.4 years?
- Not necessarily. It depends on how many observations in the sample lie right of the turnaround point. A more likely possibility is that the estimated effect of exper on wage is biased because we have controlled for no other factors, or because the functional relationship between wage and exper is not entirely corrected.

MLR: Models with Quadratics More Example

- Consider we have a fitted regression line as following:

$$
\begin{array}{lccc}
\widehat{price}= & 13.39 & -0.902\log(nox) & -0.087\log(dist) \\
 & (0.57) & (0.115) & (0.043) \\
 & -0.545rooms & +0.062rooms^2 & -0.048stratio \\
 & (0.165) & (0.013) & (0.006)
\end{array}
$$

  where $n = 506$, $R^2 = 0.603$. and

$$
\begin{aligned}
\text{nox} &= \text{nitrogen oxide in air} \\
\text{dist} &= \text{distance from employment centers, in miles} \\
\text{stratio} &= \text{student/teacher ratio}
\end{aligned}
$$

- The predicted $log(price)$ is a convex function of rooms.
- One can calculate the turnaround value of rooms is around 4.4
- However, it is hard to believe that the house price is decreasing in rooms with small number of rooms!

MLR: Models with Quadratics More Example

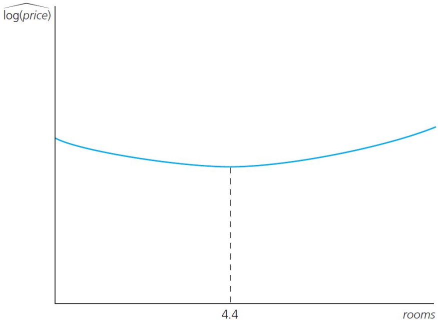

Figure: $\widehat{price}$ as a quadratic function of *rooms*

- Do we really believe that starting at three rooms and increasing to four rooms actually reduces a houses expected value?
- Probably not. It turns out that only five of the 506 communities in the sample have houses averaging 4.4 rooms or less, about 1% of the sample. This is so small that the quadratic to the left of 4.4 can, for practical purposes, be ignored.

MLR: Models with Interation Terms

- To investihate the house price, one propose a following population model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + \beta_3 sqrft * bdrms + \beta_4 bthrms + \mu$$

where $sqrft * bdrms$ is the interation term.

- Hence, the marginal effect of *bdrooms* on price is

$$\frac{\partial price}{\partial bdrms} = \beta_2 + \beta_3 sqrft.$$

- The effect of the number of bedrooms depends on the level of square footage (*sqrft*). (Why this is intuitive)?
- Interaction effects complicate interpretation of parameters: β_2 is the effect of number of bedrooms, but for a square footage of zero, which arises the interpretation difficulty.

MLR: Reparametrization of Interaction Effects

- A general model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \mu$$

can be reparametrized as

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - E(x_1))(x_2 - E(x_2) + \mu$$

where $E(x_1)$ and $E(x_2)$ are population means of x_1 and x_2, and can be replaced by their sample means.

- At home, find the relationship between $(\widehat{\alpha}_0, \widehat{\delta}_1, \widehat{\delta}_2)$ and $(\widehat{\beta}_1, \widehat{\beta}_1, \widehat{\beta}_2)$.
- Now,

$$\frac{\partial y}{\partial x_2} = \delta_2 + \beta_3 (x_1 - E(x_1)),$$

which means δ_2 is the partial effect of x_2 on y at the mean value of x_1.

- Advantages of reparametrization: 1. It is easy to interpret all parameters. 2. Standard errors for partial effects at the mean values are available (se of δ_2). If necessary, interaction may be centered at other interesting values.

MLR: Reparametrization of Interaction Effects: An Example

- Suppose one would like to investigate the effect of classes attendance on final exam, therefore, after estimation, we have following fiited regression line,

$$\begin{array}{lcccc} \widehat{stndfnl} = & 2.05 & \text{-0.0067atndrte} & \text{-1.63priGPA} & \text{-0.128ACT} \\ & (1.36) & (0.102) & (0.48) & (0.098) \\ & +0.296\text{priGPA}^2 & +0.0045\text{ACT}^2 & +0.0056\text{priGPA*atndrte} & \\ & (0.101) & (0.0022) & (0.0043) & \end{array}$$

  where $n = 680$, $R^2 = 0.229$, and

  stndful= the student final exam & atndrte is the class attendance rate
  priGPA is the prior college GPA & ACT is the standardized exam score

- However, if we simply look at the coefficient on atndrte, we will incorrectly conclude that attendance has a negative effect on final exam score. But this coefficient supposedly measures the effect when priGPA=0, which is not interesting (in this sample, the smallest prior GPA is about 0.86). (Why? we have the interact term here. The marginal effect of *atndrte* on *stnfnl* is linear in *priGPA*).
- How should we estimate the partial effect of *atndrte* on *stndfnl*? Now, since from the data, the mean of $priGPA = 0.86$, we need to rerun the regression, where we replace $priGPA * atndrte$ with $(priGPA2.59) * atndrte$.
- This gives, new coefficient on *atndrte* is 0.0078 and corresponding standard error is 0.0026, which yields $t = 3$ and implies attendance has a statistically significant positive effect on final exam score.