EXERCISE 17

WEIYU LI

The objective in this exercise is to estimate the production function for China's non-governmental businesses for the Year 2003. The data include 2052 valid observations on 7 variables:*Output, capital, Labor, Province, Ownership, Industry,* **and** *List***, where the first three variables are self-defined, and the other variables are categorical variables (for example,** *List* **= 1 if a business is a listed company, 0 otherwise). Please use the dataset (***Business03.txt***) to answer the following questions:**

(1) Run a multiple linear regression (MLR) model by regressing $\ln(Output)$ **on a constant term,** $\ln(Capital)$ **and** $\ln(Labor)$**:**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u,$$

where $Y = \ln(Output), X_1 = \ln(Capital)$**, and** $X2 = \ln(Labor)$**, and** u **is the disturbance term. Report the regression results in the standard format. That is, you need to report the regression model, the t-values, or the p-values or the corresponding standard errors for the coefficients in the model,** R^2, $\bar{R}^2$ **(Adjusted** R^2**), and the F test statistic or its corresponding p-value.**

```
library(scatterplot3d)
library(np)
data <- read.table('Business03.txt', header = T)
output <- log(data[,"output"])
log.capital <- log(data[,"capital"])
log.labor <- log(data[,"labor"])

fit <- lm(output ~ log.capital + log.labor, x = T, y = T)
# We use x,y = T here, since they are required in (4)
summary(fit) # This returns the coefficients in the regression model,
# the standard errors, t-values, and p-values for the coefficients,
# R^2, Adjusted R^2, F-test statistic and its corresponding p-value.
```

(2) Run a nonparametric regression model by regressing $\ln(Output)$ **on** $\ln(Capital)$ **and** $\ln(Labor)$**, using the local constant procedure**

$$Y = m(X_1, X_2) + u.$$

Calculate the R^2 **based upon the formula**

$$R^2 = \frac{\left[\sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})\right]^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2},$$

which is the square of the sample correlation between Y_i **and** $\hat{Y}_i$**. Plot** $\hat{Y}$ **against** X_1 **and** X_2 **in a three-dimensional diagram. Does the diagram lend any support to the MLR model in part (1)?**

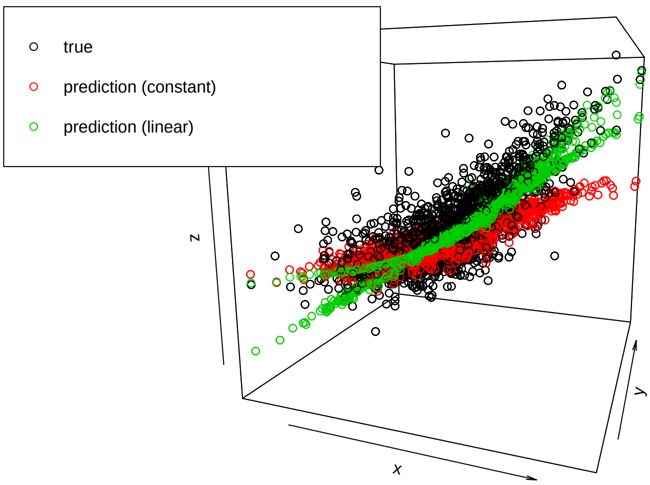
(3) Repeat part (2) by using the local linear procedure.

Date: 2019/12/09.

liweiyu@mail.ustc.edu.cn.

```
# (2)
fit.loess.c <- loess(output ~ log.capital + log.labor,
                     data.frame(output, log.capital, log.labor), degree = 0)
pred.loess.c <- predict(fit.loess.c, data.frame(log.capital, log.labor))
R.square.c <- cor(output, pred.loess.c)^2
# 0.6150078, which is R^2 calculated for (2)
scatter3D(x = log.capital, y = log.labor, output, col = 1, phi = 10, theta = 20)
scatter3D(x = log.capital, y = log.labor, pred.loess.c, col = 2, add = T)

# (3)
fit.loess.l <- loess(output ~ log.capital + log.labor,
                     data.frame(output, log.capital, log.labor), degree = 1)
pred.loess.l <- predict(fit.loess.l, data.frame(log.capital, log.labor))
R.square.l <- cor(output, pred.loess.l)^2
# 0.6645568, which is R^2 calculated for (2)
scatter3D(x = log.capital, y = log.labor, pred.loess.l, col = 3, add = T)
legend(-0.8, 0.35, c("true", "prediction (constant)", "prediction (linear)"),
       col = 1:3, pch = 1)
```



Indeed, the diagram in local constant model (the red dots) supports the linear model in (1), since it looks like a flat surface. However, we don't think local constant regression matches the data quite well and the true underlying model is linear regression.

(4) Test the correct specification of the model in part (1) using *np* package.

```
npcmstest(model = fit, xdat = data.frame(log.capital, log.labor), ydat = output)
# One can change boot.num to speed up the process
# Test result: Null of correct specification is rejected at the 0.1% level
```