Risk and Portfolio Management with Econometrics – Second Homework

Prof. Petter Kolm

Due date: Friday, October 9, 2015

Please post any questions to NYU Classes.

Questions

1. (i) Show that in the two-variable case, $R^2 = \rho_{xy}$ holds.

   (ii) Does $R^2 = \rho_{xy}$ hold in the multi-variable case? Why, or why not?

2. Consider the linear model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$$

   and assume that the classical linear regression assumptions (MLR.1-MLR.6) hold.

   (i) Show that $\dfrac{\left(\hat{\beta}_j - \beta_j\right)}{se\left(\hat{\beta}_j\right)} \sim t_{n-k-1}$

   (ii) Give an example of a joint multiple hypothesis with q linear restrictions. Show that $\dfrac{\left(SSR_r - SSR_{ur}\right)/q}{SSR_{ur}/\left(n-k-1\right)} \sim F_{q,n-k-1}$

   (Hint: Use the definitions of t- and F-distributions in the handout "A Brief Review of Probability and Statistics" from Lecture 2.)

3. Suppose that the population model determining is y

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

   and this model satisfies the classical linear regression assumptions. However, we estimate the model that omits x_3. Let $\tilde{\beta}_0$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ be the OLS estimators from the regression of y on x_1 and x_2. Show that the expected value of $\tilde{\beta}_1$ (given the values of the independent variables in the sample) is

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n} \hat{r}_{i1} x_{i3}}{\sum_{i=1}^{n} \hat{r}_{i1}^{\ 2}}$$

where the $\hat{r}_{i1}$ are the OLS residuals from the regression of x_1 on x_2. [Hint: Use that $\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$. Plug $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ into this equation.

After some algebra, take the expectation treating x_{i3} and $\hat{r}_{i1}$ as nonrandom.]

4. [Please solve this problem both in Excel and Matlab.] Use the data in WAGE2.XLS for this exercise.

(i) Consider the standard wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 \exp er + \beta_3 tenure + u.$$

State the null hypothesis that another year of general workforce experience has the same effect on log(*wage*) as another year of tenure with the current employer.

(ii) Test the null hypothesis in part (i) against a two-sided alternative, at the 5% significance level, by constructing a 95% confidence interval. What do you conclude?

5. [Please solve this problem both in Excel and in Matlab.] Use the data in DISCRIM.XLS to answer this question.

(i) Use OLS to estimate the model

$$\log(psoda) = \beta_0 + \beta_1 prpblck + \beta_2 \log(income) + \beta_3 prppov + u.$$

and report the results in the usual form. Is $\hat{\beta}_1$ statistically different from zero at the 5% level against a two-sided alternative? What about at the 1% level?

(ii) What is the correlation between log(*income*) and *prppov*? Is each variable statistically significant in any case? Report the two-sided *p*-values.

(iii) To the regression in part (i), add the variable log(*hseval*). Interpret its coefficient and report the two-sided *p*-value for $H_0: \beta_{\log(hseval)} = 0$.

(iv) In the regression in part (iii), what happens to the individual statistical significance of log(*income*) and *prppov*? Are these variables jointly significant? (Compute a *p*-value.) What do you make of your answers?

(v) Given the results of the previous regressions, which one would you report as most reliable in determining whether the racial makeup of a zipcode influences local fast-food prices?

6. The following three equations were estimated using the 1,534 observations:

$$\widehat{prate} = \underset{(0.78)}{80.29} + \underset{(0.52)}{5.44}\, mrate + \underset{(0.045)}{0.269}\, age - \underset{(0.00004)}{0.0013}\, totemp$$

$$R^2 = 0.100, \bar{R}^2 = 0.098$$

$$\widehat{prate} = \underset{(1.95)}{97.32} + \underset{(0.51)}{5.02}\, mrate + \underset{(0.044)}{0.314}\, age - \underset{(0.28)}{2.66} \log(totemp)$$

$$R^2 = 0.144, \bar{R}^2 = 0.142$$

$$\widehat{prate} = \underset{(0.78)}{80.62} + \underset{(0.52)}{5.34}\, mrate + \underset{(0.045)}{0.290}\, age - \underset{(0.00009)}{0.00043}\, totemp + \underset{(0.0000000010)}{0.0000000039}\, totemp^2$$

$$R^2 = 0.108, \bar{R}^2 = 0.106$$

Which of these three models do you prefer? Why?