Assignment 4

Regression and Forecasting Models

Professor William Greene
Phone: 212.998.0876
Office: KMC 7-90
Home page: people.stern.nyu.edu/wgreene
Email: wgreene@stern.nyu.edu
Course web page: people.stern.nyu.edu/wgreene/regression/outline.htm

Assignment 4

This exercise is based on the data set HealthData.mpj. This is a large data set (N=2039) on interviews with heads of households in Germany in the 1980s and 1990s. The variables in the data set are

- **id** = identification number
- **AGE** = age in years
- **EDUC** = education in years
- **FEMALE** = 1 if person is female, 0 if male
- **MARRIED** = 1 if person is married, 0 if not
- **HHKIDS** = 1 if there are children living in the household, 0 if not
- **INCOME** = household income (Marks/10000) (this is pre Euro days)
- **DOCVIS** = number of doctor visits in previous year
- **HOSPVIS** = number of hospital visits in previous year
- **PUBLIC** = 1 if person has public health insurance, 0 if not
- **ADDON** = 1 if person has additional insurance, 0 if not
- **DOCTOR** = 1 if DOCVIS > 0, 0 if not
- **HOSPITAL** = 1 if HOSPVIS > 0, 0 if not
- **HEALTHY** = 1 if answered 7 or higher on health satisfaction, 0-10 scale
- **HLTHSAT** = health satisfaction, 0-4 scale

You will fit various kinds of regressions with these data and examine the results and implications of the estimated models. So that we can interpret the results in percentages, begin by using Calc→Calculator to compute logIncome = loge(income) and use logIncome in the regressions below.

1. Fit a regression of logIncome on **AGE, EDUC, FEMALE, MARRIED, HHKIDS**. Report your computed results – coefficients, standard errors, $R^2$, $F$, etc. Which variables are ‘significant?’ Is the overall model $F$ statistic significant?

2. Use Calc to compute **FemaleAge = FEMALE*AGE** and **FemaleEduc = FEMALE*EDUC**. Now, add to your regression in part 1. these two new variables. Are they significant determinants of **INCOME**? What is the effect on **INCOME** of an
additional year in AGE for men? For Women? What is the effect on INCOME of an additional year of education for men? For Women?

3. Compute the square of age, Calc→Calculator then agesq = age*age. Add the square of age to the regression in part 1. What do you find? Report your results. Note, with b₁ the coefficient on age, and b₂ the coefficient on age squared, if b₁ is positive and b₂ is negative, then the profile of income (vertical axis) on age (horizontal axis) is a parabola (hill). What did you find? (You should find b₁ > 0 and b₂ < 0.) Based on your quadratic model, you can find the age at which income reaches its maximum, as implied by your model. (See pages 656-660 of your text.) According to the footnote on page 659, the derivative of (log) income with respect to age would be β₁ + 2β₂Age. At the top of the hill, this derivative would equal zero. So, you can solve for the AGE at which this is equal to zero; it will be -β₁/(2β₂). (Note, your b₂ is negative, so the ratio is positive. So, using your results in your regression, at what AGE does log Income reach its maximum. (Note, INCOME reaches its maximum at the same point where logINCOME does.)

4. Is there an income difference between male and female heads of households? The coefficient on FEMALE in the regression in part 1 will suggest an answer. What is the difference? Is it statistically significant? Now, compute the interaction variable, Calc→Calculator then FEMMAR = female*married. Add FEMMAR to the regression in Part 1. Report the results. Note, in your model, the base case for comparison is an unmarried, male head of household. What is the difference implied by your new model between an unmarried male and an unmarried female? Between an unmarried male and a married female?

5. This last exercise is concerned with a phenomenon called ‘moral hazard.’ In the presence of moral hazard, people will engage in an activity more if they are insured for it than if they are not. In your data, what this would mean is that people with insurance would visit the doctor more. Compute a regression of DOCVIS on AGE, EDUC, FEMALE, MARRIED, INCOME and PUBLIC. Report your estimated model. Which variables seem to be significant determinants of the number of doctor visits. To the question of moral hazard, does your estimated model suggest that people with health insurance (PUBLIC = 1) visit the doctor significantly more than those who do not?