Economics 402: Method of Economic Investigation
Lent Term 2009
Dr. Radha Iyengar

Exercise 1: Health and Earnings

In class we discussed the income health gradient , that is the fact that low income and poor health seem to coincide in many people. We also discussed the way in which income and health are both correlated with a range of other issues included education, occupation, age, sex, marital status and ethnicity. Below is an exercise to help you work through some of the implication of our class exercise and develop your thinking about causal inference.

To begin, suppose we wanted to estimate the effect of low income on health and we know the true function is:

$$Health_{it} = \beta_0 + \beta_1 Income_{it} + \beta_2 Education_{it} + \beta_3 Age_{it} + \varepsilon_{it} \qquad (1)$$

Question 1: DATA STRUCTURES
   A.   What type of data would you need to estimate this equation? Why?

   B.   If you had only uni-dimensional data, what estimating equation would you use?

Question 2. DATA QUALITY
You get your data and while you can perfectly measure income, you are unable to measure education and age perfectly.
A.   Algebraically, show what the bias in the OLS estimate from the equation 1 excluding education and age. In what direction do you expect this to bias your estimates?

B.   While you are not able to measure education and age perfectly, you are able to do so with some error. The error is random so that you know what you observe is:
$$\widetilde{Age} = Age + e$$
and
$$\widetilde{Education} = Education + u$$
What is the bias in your OLS estimate from using these measures?

   C.   Compare your predicted bias in A and B. Discuss which approach you would take and why.

Question 3. CAUSAL REALTIONSHIP
 For each of the factors listed above (education, occupation, age, sex, marital status and ethnicity) discuss the way in which the correlation of these issues might affect your ability to estimate the causal effect of health on education.

Question 4. DATA INTERPRETATION

Some scholars have used cross-country comparisons between the US and UK to address some of these concerns. The argument they provide is that the main excluded factor is access to medical care, which varies dramatically by education and income in the US. In the US, the NHS provides free insurance so this is not a problem.

    A.   What is the assumption such a strategy would make to "identify" the causal relationship of income on health? Does this make sense?

Below is evidence from comparison of the health gradient in the US and UK across income deciles from and IFS report.
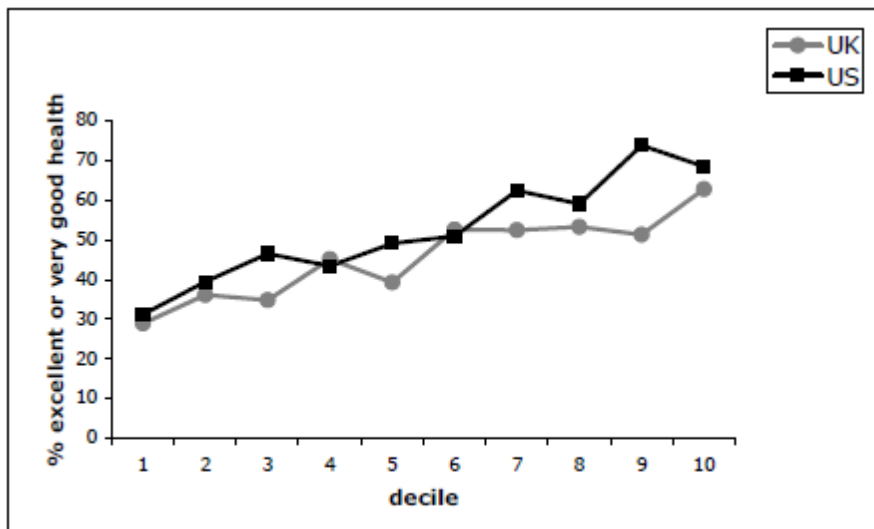

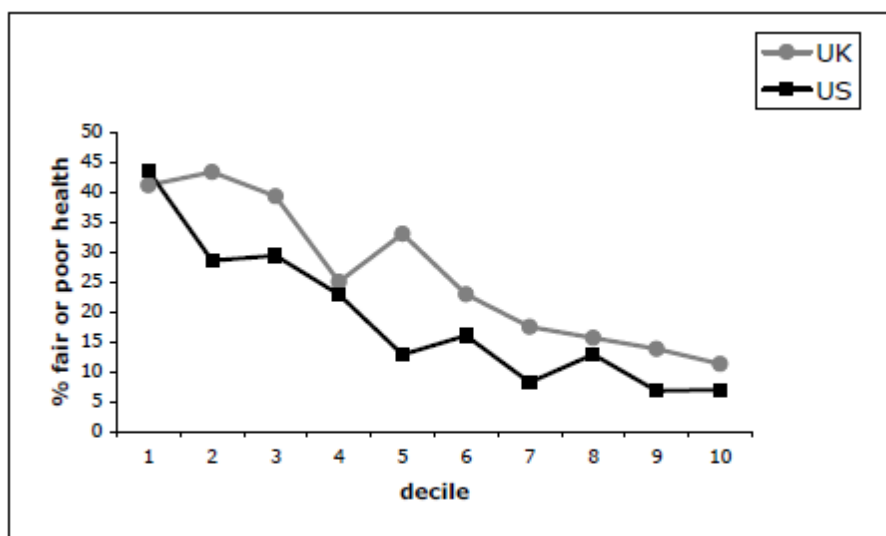Fig. 1A. % in Excellent or Very Good Health by Income Deciles.

Fig. 2A. % in Fair or Poor Health by Income Deciles.

    B.   If you believe the identifying assumption of this paper, does this evidence support a causal effect of income on health? Why or why not?
    C.   What else might explain the observed similarities and differences between the US and UK?
    D.   Can you think of a test to distinguish between the theory in A and the theory in B? Please describe.