

Supervision 2: Part IIB Paper 8
Professor Munshi

1. A public health researcher conducted a study on the effect of alcohol consumption on college grades. She discovered a negative and significant relationship between a student's grade point average (a continuous variable between 0 and 5) and the number of beers he/she drank per week. She consequently concluded that drinking had an adverse effect on college grades (this is a true story).

Suppose that you managed to get the beers-grades data from her and wanted to verify her findings. Here is the regression that you would run: $Y_i = a + bX_i + e_i$.

(a) What is X_i ? What is Y_i ? Under what statistical conditions would we get an unbiased estimate of the beer-effect? Would you expect these conditions to be satisfied in practice? If not, could you provide specific sources of bias that could arise in this case?

(b) Could you suggest ways in which the bias discussed above could be avoided?

2. An NGO launched a microcredit program in rural Bangladesh in 1990. All households owning less than $\frac{1}{2}$ an acre of land in 50 targeted villages were eligible to participate in the program. The program ran for two years and at the end of this period the NGO commissioned a research team to evaluate the impact of the program on child health and education. The research team conducted a survey of eligible households in the 50 villages and estimated the following equation:

$$y_i = \alpha + \beta C_i + \varepsilon_i$$

where y_i measures the age-adjusted nutritional level or educational attainment of the children in household i , C_i takes the value one if the eligible household participated in the program and zero if it was eligible but did not participate, ε_i is an error term.

(a) Restricting the sample to households in the 50 villages that were eligible to participate in the program, as described above, the researchers obtained a positive and significant estimate for β . They concluded that the program had a positive impact on the nutrition and the education of children in the participating households. Explain clearly why they might have been mistaken.

(b) Suppose you could re-do the analysis. With hindsight, how would you design the research study to solve the problem that you identified above?

3. Marriage in sub-Saharan Africa is exogamous; i.e. individuals must marry outside their clan. This implies that marriage in sub-Saharan Africa is accompanied by membership in a new affine network organized around the wife's family. The traditional clan-based networks have been transplanted to the city in recent decades, with urban migrants in Africa receiving support from both their birth network and their affine network (if married) in finding jobs. In turn, they send home remittances and support members of the network in the city, by providing them with accommodation, credit, and other forms of assistance. A team of researchers conducted a survey of male migrants in Kisumu, the third largest city in Kenya, in 2001. They were interested in estimating the effect of marriage on migrant incomes and so ran the following regression:

$$y_i = \alpha + \beta M_i + \gamma X_i + \varepsilon_i$$

where y_i is migrant i 's income, $M_i=1$ if he was married, 0 if he was single, X_i is the migrant's age, and ε_i is the error term. All men ultimately marry, so the regression was restricted to men aged 20-30. The researchers ran two regressions: OLS and instrumental variables. They found that β was positive in each case, but larger in magnitude with the instrumental variable regression (you can assume that the instrument is valid). When they replaced the migrant's income with the fraction of his income that he sent home to his rural network, they obtained the same pattern; β was positive, but larger in magnitude with the instrumental variable regression. Can you interpret this result by taking advantage of findings from other papers that you studied in class?