Honors General Exam

Solutions

Harvard University
April 2013

PART 3: ECONOMETRICS

Part I: Nature v. Nurture

What is the relative importance of "nature" (genes) vs. "nurture" (social and family environment) in determining economic outcomes? This part examines this question using data from a large adoption agency that placed Korean children in American families between 1964 and 1985.

At this agency, the parents must file an application, pass a criminal background check, and attend adoption classes; if all goes well, they are then deemed eligible. Children are then matched with eligible parents on a first-come, first-serve basis.

The data set contains data on the parents and their children, both adopted and non-adopted (natural), at the time of adoption and also at the end of the study when they are adults. Some households have multiple adoptees; for the purpose of this analysis, assume that the Korean adoptees in the same household are not related by blood. The analysis is restricted to adoptees who are at least 25 years of age at the end of the study.

(Figures are omitted here in the answer key, but are provided in the questions for this portion of the exam.)

Question 1

To Using regression (1) in Table 1:

(a) Compute the estimated effect on the child's years of education of a mother who drinks compared to a mother who does not drink.

Solution: Based on the coefficient on *Mother Drinks* in regression (1), the child whose mother drinks will attain 0.039 fewer years of education than the child whose mother does not drink.

(b) Compute a 95% confidence interval for your estimated effect in (a).

Solution: Given the standard error is 0.205, and we're calculating a 95% CI, the confidence interval is given by:

$$-0.039 \pm 1.96(0.205) = (-0.44, 0.36)$$

Question 2

Consider the relationship between the child's years of education and parental income, holding constant the regressors in Table 1, column (1) other than parental income.

(a) Suggest a reason why this effect might be nonlinear; that is, why should income be included in logs rather than levels.

Solution: Income tends to have the greatest marginal effect at low income levels. An extra dollar when you're very rich will have little effect on educational attainment compared to an extra dollar when you are poor.

(b) Specify both the appropriate word and the correct number: An (increase/decrease) in parental income of ___% (holding all other factors constant) has the same estimated effect on the child's years of education as having a mother who drinks compared to a mother who does not drink (holding all other factors constant; the effect from Question 1, part (a)).

Solution: Holding constant parents' education, a 100% increase in parental income results in a decrease in the years of the child's educational attainment of 0.018. Setting $-0.039 = -0.018X$ gives $x = 2.16$. Thus, an *increase* in parental income of *216%* has the same estimated effect as having a mother who drinks compared to a mother who does not drink.

Question 3

The standard errors reported in Table 1 are "clustered" standard errors, clustered at the level of the household.

(a) The author of the paper is not sure whether to use clustered or conventional heteroskedasticity-robust standard errors. Which should she use, and why?

Solution: The author should use clustered standard errors. This allows the unobserved determinants of education to vary by family.

(b) How will the clustered standard errors reported in Table 1 compare in size to conventional heteroskedasticity-robust standard errors? Explain.

Solution: We cannot tell without running the regression. They could be larger if there is positive correlation between error terms within families or smaller if there is negative correlation between individuals within families.

Question 4

Consider a female adoptee whose adoptive mother and father both have 16 years of education, whose parents' income is $100,000, mother's BMI is 23, father's BMI is 24, the mother does not drink, and the father does not drink. Also suppose that the child was adopted in the initial program year (so all binary year variables equal zero).

(a) Using regression (1) in Table 2, compute the probability that the child graduates from college.

Solution: Use the given characteristics and the regression coefficients from the regression to calculate the z-value:

$$= 0.142 + 0.057(Mother's\ Education) - 0.010(Father's\ Education) + 0.008\log(Parents'Income) - 0.086(Mother's\ BMI) - 0.003(Father's\ BMI)$$

$$= 0.142 + 0.057(16) - 0.010(16) + 0.006\log(100{,}000) - 0.086(23) - 0.003(24)$$

$$= -1.064$$

For the given individual, the z-value is -1.064, and $\Phi(-1.064) = 14.5\%$. (You'll need a z-score table to determine this. You'll be given one on the exam if one is necessary to solve the problem. Here, because the z-value is negative, you'd look on the table for $1 - \Phi(1.064) = 1 - 0.8554 = 0.145 = 14.5\%$.)

(b) What is the difference in the predicted probabilities of graduating from college for the adoptee in (a) compared with a male adoptee with all other characteristics identical?

Solution: The z-value for the male student is -1.46, so his probability of his graduation from college is $\Phi(-1.46) = 1 - \Phi(1.46) = 1 - 0.9279 = 7.2\%$. Given that there is a 14.5% chance that the female student in (a) will graduate from college, she is 7.3 percentage points more likely to graduate than the male student.

(c) Now use the linear probability model from Table 1, regression 3 to estimate the change in predicted probabilities for the comparison in Question 4, part (b) (that is, male vs. a female adoptee, with the values of the other regressors given at the beginning of this question).

Solution: Because Table 1 reports OLS estimates, the change in probability is given by the coefficient on "Child is male." According to regression 3, the male child has a 15.9 percentage point lower probability of graduating from college than the female student.

Part II: Fast-Food TV Advertising and Childhood Obesity

Childhood obesity is a health problem of significant concern. In the 1960s, approximately 4 percent of American children ages 6 to 11 were overweight; by 1999, 13 percent of American children were overweight. Measured in terms of BMI, the average BMI for children rose from 16.63 in the 1960s to 17.37 in 1999, an increase of almost 5%; this is a large increase in historical and medical terms. [The BMI is the body mass index, which is weight (in kilograms) divided by the square of height (in meters), so the units of the BMI are kg/m^2.]

A shift to a high-fat, high-calorie childhood diet – the sort of food found at fast-food restaurants – is one possible reason for the increase in childhood BMI. This section considers whether exposure to fast-food advertising on TV plays a role in this increase.

The data set is a cross-sectional data set on children aged 6-11 in the U.S. in 1997. It contains data on children's characteristics, family characteristics, TV viewing by the child, and characteristics of the child's county.

Question 1

Suggest a reason why TV exposure might be endogenous in regression (1).

Solution: Kids with a higher BMI may spend more time watching TV and thus may see more hours of television ads per week. Therefore, TV exposure may be an endogenous variable.

Question 2

The Regression (3) instruments for TV exposure using three variables: the price of TV advertising, the number of households with TV, and temperature. Assume for this question that these instruments are exogenous.

(a) Suppose the instruments used in regression (3) are weak. How will the estimated coefficient compare to (i) the OLS estimate from regression (1) and (ii) the true value of the coefficient?

Solution: If the instruments are weak, the TSLS estimates will be biased towards the OLS estimates. Thus, the estimated coefficient will lie between the true value and the OLS estimate.

(b) Is there evidence in Table 3 that this set of instruments is weak? Explain.

Solution: The first stage F-statistic for the three instruments is 42, which is greater than the cutoff of 10. (An F-stat of greater than 10 indicates that an instrument is strong.) Thus there is no evidence in this table that these instruments are weak.

Question 3

After seeing the results of your study, Mayor of New York City Michael Bloomberg calls you and tells you he's decided to ban all televisions in order to reduce childhood obesity in NYC. Is this a reasonable conclusion given the results of Table 3? Explain.

Solution: This is not necessarily a valid conclusion. There may be omitted variables that are correlated both with the instruments and the BMI of children. Three examples are population density (correlated with the price of advertising and BMI), religious beliefs (which may affect whether a family owns a TV and also participation in physical activity or diet), and region of the country (which may be correlated both with eating habits and temperature).

Question 4

To control for unobserved local characteristics, the author of the paper considers including county-level fixed effects in regression (3). What will be the effect on the first stage in regression (2)?

Solution: The county fixed effects are collinear with the instruments since they vary at the county. Thus there will be perfect multicollinearity and the first stage will not be identified.

Question 5

Your thesis advisor suggests including the number of households with TV and temperature as controls rather than instruments in all three regressions in Table 3, leaving the price of TV advertising as the only instrument for TV exposure.

(a) Is the price of TV advertising an exogenous instrument in this new regression? Explain.

Solution: This is no better than the regression in Table 3 as there still may be omitted variables. Population density is an example that may still impact both BMI and the price of advertising even when controlling for temperature and the number of households with TV.

(b) Given a choice between regression 3 and the specification described here, which would you prefer? Why?

Solution: Given that temperature and the number of TVs per household are unlikely to be exogenous instruments, they are better used as controls. Thus, given the choice, I would prefer the regression described in this question.