

Exam #2 Answer Key

Economics 435: Quantitative Methods

Spring 2006

1 Applications

There are a wide variety of correct answers, so your answer does not need to look like mine in order to be correct.

By the way, I made our hypothetical data set into a panel data set primarily as a distraction. This particular set of questions would probably best be answered using just the cross-section. The reason for this is that we are probably interested in how economics education affects savings rates across a person's entire working life. The effect of *ECON* would be identified in a fixed effects analysis entirely from people whose value changed between 2005 and 2006, i.e., currently-enrolled university students.

a) My dependent variable would be the savings rate, i.e. $SAVRATE \equiv SAVINGS/INCOME$. My explanatory variables would be $(AGE, AGE^2, SEX, UNIV, ECON, \ln INCOME)$.

The tricky part here is that savings rates will behave very strangely for individuals with very low or even negative income. For example, someone with zero income necessarily has zero or negative savings, implying a savings rate of either 0/0 (undefined) or $-\infty$. What I would do is narrow my question to the effect of economics education on savings rates among individuals with incomes over (for example) \$40,000 per year. That is, I will simply drop everyone from my sample with income below that level. I would also drop everyone from my sample outside of the age range 25-64.

The next issue is that the savings rate has an upper bound of 1 (it has no lower bound; you can earn \$40,000 in a year and spend \$10 billion if you have \$10 billion in wealth at the start of the year). So we will estimate a Tobit model with no lower bound and an upper bound of 1.

The outcome of interest would be y or *SAVERATE*.

b) My dependent variable would be the total RRSP contribution, i.e. *RRSP*. My explanatory variables would be $(AGE, AGE^2, SEX, UNIV, ECON, \ln INCOME)$.

There is clearly an upper bound and a lower bound on RRSP contributions, so that suggests a Tobit model. Unfortunately, while we know the lower bound (zero) the upper bound depends on information (previous year's RRSP contributions and limits, as well as how much of the person's income is "earned" by Revenue Canada's definition) that we simply do not have. We have three options I can think of:

- Ignore the upper bound. Estimate a Tobit model with the lower bound of zero and no upper bound.
- Redefine the outcome to a binary indicator of whether the respondent contributed at least (for example) \$2,000 to his or her RRSP. Estimate a probit or logit model for this outcome.
- Top-code the contribution. Define a new variable

$$RRSP2 = \min(RRSP, 10000)$$

and estimate a Tobit model with *RRSP2* the dependent variable, a lower bound of zero, and an upper bound of 10,000.

In options 1 and 2 the relevant outcome is y . In option 3, it is hard to say what the relevant outcome is since we have a combination of corner solution outcomes (outcomes at the lower bound) and censored outcomes (outcomes above the upper bound).

c) Here the clear problem is that we observe asset allocation only for those respondents who have assets to allocate. Thus we have to decide which of these questions is the one we are interested in answering:

1. How do economics courses influence the asset allocation decisions of those who have assets to allocate?
2. How do economics courses influence the asset allocation decisions that a person would make if he/she had assets to allocate?

If we want to answer question 1 above, we can simply restrict the sample to those who made positive RRSP contributions in the previous year. The dependent variable would be *STOCKS*, and the explanatory variables would be (*AGE*, *AGE*², *SEX*, *UNIV*, *ECON*, \ln *INCOME*). We may also want to include *RRSP* as an explanatory variable but whether that is a good idea or not is hard to say. The model would be a Tobit with an upper bound of one and a lower bound of zero.

If we want to answer question 2 above, we need to set up a model in which everyone makes a hypothetical asset allocation decision but that decision is only observed if the person makes an RRSP contribution. This is a standard incidental truncation problem. I would set up a Heckman-style model, in which the selection variable is:

$$s \equiv I(RRSP > 0)$$

The selection model would be a probit with (*AGE*, *AGE*², *SEX*, *UNIV*, *ECON*, \ln *INCOME*) as the explanatory variables. I would then estimate a Tobit model with *STOCKS* as the dependent variable and with (*AGE*, *AGE*², *SEX*, *UNIV*, *ECON*) and the Heckman correction term based on the selection model as the explanatory variables. The upper and lower bounds of the Tobit model would again be zero and one respectively.

In either of these cases, the relevant outcome would be *STOCKS*.

2 A variation on the Heckman correction

a) This CEF is identified in our data: we have a random sample that includes all of its variables so we can identify any feature of their joint distribution.

b) We have:

$$E(y|x, z, s) = \beta_0 + \beta_1 x + \beta_2 s + \rho \lambda(x, z, s)$$

c) You can follow a two-step procedure:

1. Estimate the γ coefficients using standard maximum likelihood techniques for estimating a probit model.
2. Estimate an OLS regression of y on $x, s, \hat{\lambda}$, where

$$\hat{\lambda}_i \equiv s_i \frac{\phi(\hat{\gamma}_0 + \hat{\gamma}_1 x_i + \hat{\gamma}_2 z_i)}{\Phi(\hat{\gamma}_0 + \hat{\gamma}_1 x_i + \hat{\gamma}_2 z_i)} - (1 - s_i) \frac{\phi(\hat{\gamma}_0 + \hat{\gamma}_1 x_i + \hat{\gamma}_2 z_i)}{1 - \Phi(\hat{\gamma}_0 + \hat{\gamma}_1 x_i + \hat{\gamma}_2 z_i)}$$

It is also possible to estimate this model by maximum likelihood.