Econometrics II

Prof. Brendan Kline

Homework on Discrete Outcomes Models

and also maximum likelihood estimation

Due on October 4 (in class)

This homework asks you to consider some questions relating to models with discrete outcomes:

1. Sometimes people "translate" the estimated coefficients between different binary outcomes models (e.g., between a linear probability model and a probit model) estimated on the same dataset. To see how, consider the setup of a general linear index binary outcomes model:

$$P(y = 1|x) = G(x\beta)$$

where $G(\cdot)$ is a known function. In the linear probability model, $G(z) = z$. In the probit model, $G(z) = \Phi(z)$, the standard Normal CDF. In the logit model, $G(z) = \frac{e^z}{1+e^z}$, the standard logistic CDF. Denote the parameters as β_{LPM}, β_{probit}, and β_{logit}. You can think of these parameters as coming from estimating these three models, all based on the same dataset.

By definition, the marginal effect of explanatory variable k (i.e., x_k) is $\frac{\partial P(y=1|x)}{\partial x_k}$.

   (a) What is the simplified expression for the marginal effect of x_k in the linear probability model? In the probit model? In the logit model? (plug in the respective G functions to the general formula and take the derivative)

   (b) The marginal effects in the probit model and the logit model are functions of x. Answer the following questions for the probit model, and separately for the logit model. The answers will depend on the value of β in that model.

i. Give a simple characterization of the entire set of values of x that solve $\max_x \left| \frac{\partial P(y=1|x)}{\partial x_k} \right|$. In other words, characterize the values of x at which the marginal effect of x_k has greatest magnitude.

ii. What is $\max_x \left| \frac{\partial P(y=1|x)}{\partial x_k} \right|$? In other words, what is the absolute value of the marginal effect of x_k when it has the greatest magnitude (where "greatest" means as a function of x)?

iii. What is the marginal effect of x_k (i.e., $\frac{\partial P(y=1|x)}{\partial x_k}$) when $x = 0$? If the answer depends on any "complicated" expressions (e.g., the exponential or the square root of a number), then also give a simple numerical approximation of the form $c\beta_k$ where c is a specific number.

(c) Now, it is possible to "relate" the parameters across models. The idea is that the marginal effect of x_k should be "the same" across all model specifications (i.e., LPM, probit, logit). Therefore, we should "expect" that the relationship between β_{LPM}, β_{probit}, and β_{logit} is such that indeed the marginal effects of x_k across all model specifications are "the same." The relationships you derive below are the standard ways of translating the β coefficients between the three models, but should be used with caution because (as now seen) they are only heuristic approximations.

i. Suppose that the marginal effect of x_k in the linear probability model is equal to the marginal effect of x_k in the probit model, at $x = 0$. What would that imply about the relationship between β_{LPM} and β_{probit}?

ii. Suppose that the marginal effect of x_k in the linear probability model is equal to the marginal effect of x_k in the logit model, at $x = 0$. What would that imply about the relationship between β_{LPM} and β_{logit}?

iii. Suppose that the marginal effect of x_k in the probit model is equal to the marginal effect of x_k in the logit model, at $x = 0$. What would that imply about the relationship between β_{probit} and β_{logit}?

2. Another common model for discrete[1] outcomes is the Poisson model. According to the Poisson model, the outcome Y can be any non-negative integer (i.e., $\{0, 1, 2, \ldots\}$). Specifically,

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where $\lambda > 0$ is the unknown parameter.

(a) Prove that the expectation of a Poisson random variable with parameter λ is λ. You can do this by using the definition of expectation (plugging in the expression for $P(Y = y)$). Hint: start the infinite sum at $y = 1$ since $y = 0$ does not contribute to the expectation. Also, you will need to use the fact that $\sum_{s=0}^{\infty} \frac{\lambda^s}{s!} = e^{\lambda}$, so try to rearrange the expression for the expectation so that $\sum_{s=0}^{\infty} \frac{\lambda^s}{s!}$ appears.

Also, you do not need to prove it, but another useful fact for later questions is that the variance of a Poisson random variable with parameter λ is also λ. And therefore the expectation of Y^2 is $\lambda + \lambda^2$.

(b) What is the log likelihood of an observation Y_i? What is the log likelihood of a random sample $\{Y_i\}_{i=1}^{N}$?

(c) What is the maximum likelihood estimator $\hat{\lambda}_N$ of λ? Does this make sense given the answer to part 2a?

(d) Evaluate the matrix J in the statement of theorem 3.3 in Newey-McFadden to find the asymptotic distribution of the MLE estimator. Assume the other conditions of theorem 3.3 hold (do not prove them)! What is the asymptotic distribution of $\sqrt{N}(\hat{\lambda}_N - \lambda_0)$, where λ_0 is the true value of the parameter?

(e) Apply a central limit theorem directly to the formula for the MLE estimator from part 2c. According to a central limit theorem, what is the asymptotic distribution of $\sqrt{N}(\hat{\lambda}_N - \lambda_0)$? You can assume the conditions of a central limit theorem apply, just

[1] Technically, the model is discrete but not finite. It is not finite because there are infinitely many options for the Y variable. However, it technically still qualifies as discrete, essentially because there are discrete "gaps" between the options for the Y variable. Also, there are only countably many options for the Y variable, rather than uncountably many as in the typical continuous outcome (e.g., where Y can be any positive real number.)

figure out what the covariance of the limiting Normal distribution will be. (Hint: assume that theorem 3.2 of Wooldridge applies, what is B?) How does this answer compare to the answer in part 2d?

3. This question is about using Stata to estimate various discrete outcomes models. Use the data in `bwght.dta`. This is cross-sectional data on birth records. Each observation is one birth. The data includes these variables for each birth: `cigs` (the number of cigarettes the mother smoked per day – on average – while pregnant), `motheduc` (mother's education in years), `white` (a binary variable indicating white race), and `lfaminc` (the logarithm of family income). Use the command `summarize cigs motheduc white lfaminc` to see the summary statistics for these variables. (Do not submit the results with your answers, rather just do so to get a sense of the data!)

   Along with your written answers to the questions, also provide a printout of your Stata log file.[2] Indicate (using pen/pencil) which parts of your Stata log correspond to the various questions asked below.

   (a) Use the command `generate smokes = (cigs > 0)`. This will generate a new binary variable that indicates whether the mother smoked at all during pregnancy. Using the `summarize` command, what fraction of mothers in the dataset smoked while pregnant?

   (b) The following parts all concern estimating the effect of education on smoking while pregnant.

      i. Estimate a linear probability model relating `smokes` to `motheduc`, `white`, and `lfaminc`. Use the `regress` command to do so. Use the command `help regress` to learn how to use the `regress` command, if needed. What is the effect of one additional year of mother's education on the probability the mother smokes while pregnant?

[2] Start the Stata log by using the File - Log - Begin... option, and conclude the Stata log by using the File - Log - Close option. Then use File - Log - View... to view the completed log file, and print.

ii. Using the `probit` command, estimate a probit model relating `smokes` to `motheduc`, `white`, and `lfaminc`. Use the command `help probit` to learn how to use the `probit` command, if needed. It is very similar to using the `regress` command. Then answer these questions relating to the results:

A. Using the actual formula for the marginal effect of a discrete explanatory variable (i.e., equation 15.15 in Wooldridge where G is specified appropriately), what is the effect of one additional year of mother's education on the probability the mother smokes while pregnant? Remember that marginal effects in non-linear models like the probit model depend on the "starting value" of the explanatory variables. Evaluate the effect where the "starting value" is the averages of all of the explanatory variables.

B. Using the "standard" derivative formula for the marginal effect of an explanatory variable (i.e., equation 15.13 in Wooldridge where G is specified appropriately), what is the effect of one additional year of mother's education on the probability the mother smokes while pregnant? Note that the "standard" formula technically is valid only for continuous variables, but it can be used mechanically for any sort of variable. As before, evaluate the effect where the "starting value" is the averages of all of the explanatory variables.

C. Use the command `mfx` after estimating the model[3]. This command gives you the marginal effects, in the sense that Stata basically applies the formulas you used by hand in the earlier parts. What is the effect of one additional year of mother's education on the probability the mother smokes while pregnant? As before, evaluate the effect where the "starting value" is the averages of all of the explanatory variable. Use the command `help mfx` to learn how to use the `mfx` command, if needed. (What does Stata do by

[3]Or, you can use the command `margins`, but it is more complicated. You may find reading examples 1-3 and example 17 of the help file for `margins` useful, available at *http://www.stata.com/manuals13/rmargins.pdf.*

default in terms of the "starting value"?)

iii. Using the `logit` command, estimate a logit model relating `smokes` to `motheduc`, `white`, and `lfaminc`. Using the command `mfx` after estimating the model, what is the effect of one additional year of mother's education on the probability the mother smokes while pregnant? As before, evaluate the effect where the "starting value" is the averages of all of the explanatory variables.

iv. Compare the estimates of the marginal effect of education on smoking between the linear probability model, probit model, and logit model. How do the estimates relate?

   Remember to submit your Stata log file!

v. NOT TO BE TURNED IN: Also, if you are interested, you can check out how the estimated β coefficients (not the marginal effects!) relate to each other across the linear probability model, probit model, and logit model, compared to the "heuristic" relationships derived in question 1 above. The actual relationships might not follow the "heuristic" relationships!