Chapter 11
Qualitative Choice Models

1 Introduction

We have seen how to transform qualitative information into dummy variables so that the qualitative information can be used as explanatory variables to help improve the fit/prediction of our regression model. We are now going to extend the analysis and allow qualitative information to be the independent variable in the regression model. We will focus on qualitative information where there are only two possible classes of observations (male and female, union worker and non-union worker, did you vote yes or no, etc.) as opposed to qualitative information where there are more than two possible classes of observations (did you vote for Bush, Gore, Nader, Buchanan, etc.). When we only have two possible classes of observations we have binary choice models. Our goal will be to determine the probability that an individual with a given set of characteristics will choose one choice rather than the other.

2 Binary Choice Models

Binary choice models are models in which the dependent variable takes on only the values 0 and 1. Again, our goal is to determine the probability with which a given individual falls into one category or the other. We will discuss two types of binary choice models, the linear probability model and the probit model.

2.1 Linear probability model

The linear probability model assumes the relationship between the binary dependent variable and the independent variables is linear. Thus, $Y_i = \alpha + \beta X_i + \varepsilon_i$. This is nothing new to us, the only difference being that now Y_i only takes on the values 0 and 1.

We can also look at this in terms of probabilities. Let the $\Pr(Y_i = 1) = P_i$. Then, $\Pr(Y_i = 0) = (1 - P_i)$ (since Y_i is either 0 or 1). So, $E[Y_i] = 1 * (P_i) + 0 * (1 - P_i) = P_i$. Also, from our regression model, $E[Y_i] = E[Y_i = \alpha + \beta X_i + \varepsilon_i] = \alpha + \beta X_i$. Putting the two together gives us: $P_i = \alpha + \beta X_i$. However, this is only true IF $0 \leq \alpha + \beta X_i \leq 1$, since probabilities must be between 0 and 1. If $\alpha + \beta X_i \leq 0$, then $E[Y_i] = 0$. If $\alpha + \beta X_i \geq 1$, then $E[Y_i] = 1$. When we estimate the linear probability model we use ordinary least squares (just run a regular regression in SAS, specifying a binary variable as the dependent variable in the model statement in SAS), and we can interpet the coefficient estimates for the linear probability model as the effect of a one-unit change in the independent variable on the probability of the dependent variable. For instance, if our coefficient estimate for β_2 is $\beta_2 = .10$, with the linear probability model we can say that a one-unit increase in X_2 will increase the probability of the observation being a 1 by 10%. NOTE: This is only true for the linear probability model, and not the probit model.

2.1.1 Problems with linear probability

One of the main problems with using the linear probability model – for observations close to the extremes of the observation interval for each of the independent variables we may have a predicted probability that is greater than 1 or less than 0, which is not possible in practice. So we arbitrarily assign the predicted values a value of 1 if $\widehat{\alpha} + \widehat{\beta} X_i \geq 1$ and 0 if $\widehat{\alpha} + \widehat{\beta} X_i \leq 0$. However, we may be assigning an "incorrect" probability that an event will occur, as in each case we are assigning a 100% chance that an observation with a given set of individual characteristics will be either a 1 or a 0.

A second problem that occurs is a statistical problem – the linear probability model is heteroscedastic. Fortunately we know the form of the heteroscedasticity – $\hat{\sigma}_i^2 = \hat{Y}_i[1 - \hat{Y}_i]$. Thus we can use weighted least squares to find the efficient estimates.

The first problem is a much bigger problem than the second, and the probit model was developed to insure that we do not have predicted probabilities greater than 1 or less than 0.

3 Probit model

The probit model is the only non-linear model we will discuss in this class. Since it is non-linear we CANNOT use the least squares procedure to estimate the coefficients, but we will use the method of maximum likelihood. Maximum likelihood is a fairly intuitive concept, although the estimation of the model can be messy. Basically, maximum likelihood estimators "look" at your data and determine what parameters (recall that the parameters are the unknown coefficient estimates, β_1, β_2, β_3, etc.) could have generated the data that you have observed.

3.1 Technical details of maximum likelihood

Technically the maximum likelihood estimator is found by maximizing the likelihood function (hence the name *maximum likelihood*). The likelihood function is just the product of the marginal density functions of the observations. More or less, a density function just tells us the probability that a given observation will occur (this is a true statement for discrete density functions, not as true for continuous density functions but true enough for our purposes). So we just multiply the marginal densities for each observation together and then maximize the likelihood function (this involves taking the derivative of the density function with respect to the unknown parameters, setting them equal to zero, and solving for the unknown parameters). Generally we cannot find analytical solutions to the likelihood function because it is non-linear so we must use an estimation routine to find these parameters (an iterative routine like we used for the Cochrane-Orcutt estimation of ρ in chapter 6). I know that I cautioned you that we could not be certain that we found the minimizing error sum of squares when we used the Cochrane-Orcutt method to estimate ρ, but you can rest assured that we WILL move towards the global maximum of the likelihood function, and find the true maximum, at least in the case of the Probit model.

3.2 The model

We have been writing all of our regression models as linear functions of the parameters, $Y = \beta_1+\beta_2X_2+\beta_3X_3+\varepsilon$. The probit model is a non-linear function of the parameters. If we let P_i be the probability $Y_i = 1$, then $P_i = F(\alpha+\beta X_i)$, where $F(\bullet)$ represents some cumulative probability distribution. The probit model assumes that $F(\bullet)$ is the normal probability distribution (you should see the notes on probability distributions on the web site). The normal probability distribution is a mess:

$$P_i = F(Z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} e^{-\frac{1}{2}t^2} dt$$

This tells us the probability that a given value will be less than Z_i. Since we will have $Z_i = \alpha + \beta X_i$, we can substitute $\alpha + \beta X_i$ in for Z_i in the equation above. Now, remember that we need to multiply all of these marginal density functions together, and THEN take the derivative. The resulting function will not be a pretty thing to take the derivative of, which is why we won't actually do the maximization.

3.3 Interpretation

Estimating the model with SAS is quite simple, as it only involves one more line of code than the ordinary least squares regression model. The output file will look very similar to the ordinary least squares regression output – you will have coefficient estimates and standard errors, and from there you can construct t-statistics. However, interpretation of the coefficient estimates of a probit model are a little difficult to understand. The actual coefficient estimate tells us the marginal effect (one-unit change) of the independent variable on the underlying Z variable. The Z variable is really a continuous variable that represents different degrees along a spectrum. This is probably best explained with an example. Suppose we are trying to determine whether a person will vote Republican or Democrat. We have all of our data and we let $vote = 1$ if this person voted Republican and $vote = 0$ if this person voted Democrat. We then have independent variables (income, age, schooling, etc.) that we use to predict how a person would vote. We get estimated coefficients for these variables – these coefficients tell us the marginal effect on the underlying Z variable, NOT on the probability of voting Republican or Democrat. In this particular example, the underlying Z variable may be viewed as a range of political views, where the range moves from liberal to conservative. Thus, the estimated coefficient tells us how much more liberal or conservative a person will come if we add an extra unit of income, or another year of age, or another year of schooling, etc. This is slightly different than what the estimated coefficient of the linear probability model tells us – it tells us the change in probability associated with a one-unit change in the independent variable.

One useful thing to look at is the sign of the variable, as it tells us which direction we are moving in along the range.

3.4 Statistical testing

Statistical testing of probit estimates is simple. The maximum likelihood estimator is approximately distributed as a normal random variable for large samples, which means that we can use either the normal distribution or the t-distribution for hypothesis testing. I suggest using the t-distribution since this is what we have been using all along. Recall that to perform a hypothesis test we take: $\frac{\hat{\beta}-\beta_{null}}{s_{\hat{\beta}}}$, which turns into $\frac{\hat{\beta}}{s_{\hat{\beta}}}$ if we are testing the null hypothesis that $\beta = 0$.

3.5 Prediction

There is one more point I want to make with the probit model. We have been predicting the value of the dependent variable based on the estimated regression coefficients and the observed independent variables throughout the course. We cannot do this as easily with the probit model since it is a non-linear model. You cannot just plug in the X variables, add them up, and then assume that they equal the Y variable.