Models of Qualitative Binary Response

Probit and Logit Models

October 6, 2015

Dependent Variable as a Binary Outcome

Suppose we observe an economic choice that is a binary signal. The focus on the course is to estimate parameters to test economic theories and/or predict the impact of exogenous change due to policy.

An important starting point is the idea of a conceptual choice model. There are two primary classes of models:

- Random Utility or Discrete Choice Model: individual examines how $\mathbf{x}$'s vary across the state of the world if 'yes' versus 'no'.
- Varying Parameters Model: State of the world doesn't vary over the 'yes' and 'no', but the individual's behavioral parameters vary if yes versus no.

The Random Utility Model (RUM)

Imagine observing a binary signal of the form [Yes,No] or [True,False]. By convention, we will always code our dependent variable $d_i = 1$ if Yes or True and $d_i = 0$ if No or False.

In the RUM, an individual i looks at the indirect utility (an index of well-being) of yes and no:

$$
\begin{aligned}
d_i = & 1 \quad if \quad V(\mathbf{X}_{i,Yes}, \epsilon_{yes}|\beta) > V(\mathbf{X}_{i,No}, \epsilon_{No}|\beta) \\
& 0 \quad if \quad V(\mathbf{X}_{i,No}, \epsilon_{No}|\beta) > V(\mathbf{X}_{i,Yes}, \epsilon_{Yes}|\beta)
\end{aligned}
$$

I will refer to the 'Yes' and 'No' conditions, the choice alternatives.

The Random Utility Model (RUM): An example

Suppose we observe potential lottery ticket buyer for a lottery ticket costing c. The individual has income Y_i and if she purchases will receive expected payouts P, and if not will receive expected payout 0. Assume that there are other factors, Z_i that are individual characteristics that are observed such as age and other demographic information.

For simplicity, suppose the individual's indirect utility is linear in parameters:

$$\begin{aligned} V(\mathbf{X}_{i,Yes}, \epsilon_{i,yes}|\beta) =& \beta_Y(Y_i - c) + \beta_P P + \beta_Z Z_i + \epsilon_{i,Yes} \\ V(\mathbf{X}_{i,No}, \epsilon_{i,No}|\beta) =& \beta_Y(Y_i - 0) + \beta_P 0 + \beta_Z Z_i + \epsilon_{i,No} \end{aligned}$$

Simplify this to find the condition of $V(Yes) > V(No)$.

Characterizing the Choice

So, we observe the individual voting yes ($d_i = 1$) iff

$$\beta_Y(Y_i - c) + \beta_R R + \beta_Z Z_i + \epsilon_{i,Yes} > \\ \beta_Y(Y_i - 0) + \beta_R 0 + \beta_Z Z_i + \epsilon_{i,No}$$

$$\beta_Y(Y_i - c) - \beta_Y(Y_i - 0) + \beta_R R - \beta_R 0 + \\ \beta_Z Z_i - \beta_Z Z_i > \epsilon_{i,No} - \epsilon_{i,Yes}$$

$$-\beta_Y c + \beta_R R > \epsilon_{i,No} - \epsilon_{i,Yes} \tag{1}$$

Note: variables that do not vary over the choice alternatives drop out of the difference, given our linear function.

The Varying Parameters Model

The varying parameters departs from the RUM in an important way. Here, variation in individual attributes is assumed to be the important determinant driving the individual's decision to purchase or not.

Allow parameters to vary over the choice alternative

Suppose instead, we focus on the vector of socio-demographic characteristics $\mathbf{Z}_i$. The individual's choice of purchasing a ticket depends on those characteristics alone. Suppose there are two characteristics in $\mathbf{Z}_i$: age (A_i) and income (Y_i). An individual would vote 'Yes' iff

$$\beta_{Yes}^{age} A_i + \beta_{Yes}^{Y} Y_i + \epsilon_{i,Yes} > \beta_{No}^{age} A_i + \beta_{No}^{Y} Y_i + \epsilon_{i,No}$$

Varying Parameters, cont.

In the binary case we consider here, the individual purchases if

$$(\beta^{age}_{Yes} - \beta^{age}_{No})A_i + (\beta^{Y}_{Yes} - \beta^{Y}_{No})Y_i > \epsilon_{i,No} - \epsilon_{i,Yes}$$

But when we estimate the model, we can't identify all these parameters, rather only:

$$\beta^{age}A_i + \beta^{Y}Y_i > \epsilon_{i,No} - \epsilon_{i,Yes}$$

where $\beta^{age} = \beta^{age}_{Yes} - \beta^{age}_{No}$ and $\beta_Y = \beta^{Y}_{Yes} - \beta^{Y}_{No}$. If there are J choice alternatives, we will recover $J-1$ sets of the K parameters- they are all normalized on one choice alternative.

Econometrics Step 1

As researchers, we can observe $\mathbf{Z_i}$ and/or $\mathbf{X_{ik}}$ (for all choice alternatives), but we can't observe the ϵ's. Nor is there a straightforward way to construct estimated errors as in OLS. Since the binary variable signals if the indirect utility is higher or not, not the degree to which it is higher. But we can tackle this problem in a maximum likelihood framework. In a RUM context, write the probability that individual i choose 'Yes' as

$$Prob(Yes|\beta, \epsilon, \mathbf{X}_{Yes}, \mathbf{X}_{No}, Z_i) = Prob(-\beta_Y c + \beta_P P > \epsilon_{i,No} - \epsilon_{i,Yes}) \tag{2}$$

Or, we can write a similar expression in a Varying Parameter Context:

$$Prob(Yes|\beta, \epsilon, \mathbf{X}_{Yes}, \mathbf{X}_{No}, Z_i) = Prob(\beta^{age} A_i + \beta^Y Y_i > \epsilon_{i,No} - \epsilon_{i,Yes}) \tag{3}$$

Note: the remainder of this presentation only presents the RUM model, but the results can easily be extended to the varying parameters framework.

Econometrics Step 2

Following Greene rewrite our RUM condition,

$$Prob(-\beta_Y c + \beta_P P > \epsilon_{i,No} - \epsilon_{i,Yes})$$

$$Prob(\mathbf{x}_i \beta > \epsilon_{i,No} - \epsilon_{i,Yes}) \tag{4}$$

So we need a probability model that assigns low probability if $\mathbf{x_i}\beta$ is small and higher probabilities if $\mathbf{x_i}\beta$ is large. Formally, we want

$$\lim_{\mathbf{x}_i\beta\to+\infty} Prob(d_i = 1|\mathbf{x}_i\beta) = 1$$
$$\lim_{\mathbf{x}_i\beta\to-\infty} Prob(d_i = 1|\mathbf{x}_i\beta) = 0$$

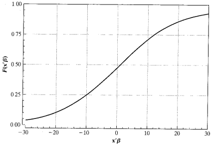

Figure: Plot of CDF

Econometrics Step 3

Since by definition, a probability is bounded by [0,1] we can use maximum likelihood estimation to recover the estimates for β. Common practice is to assume that the errors are i.i.d. Normal (Probit) or Logistic (Logit) distributed.

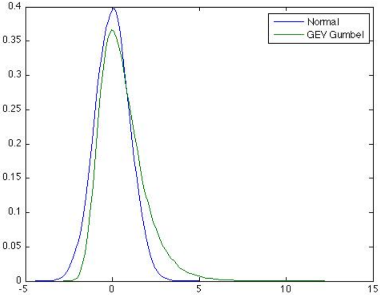

Figure: Plots of the pdf

The Likelihood Function

$$\text{Probit} : Prob(d_i = 1|\mathbf{x}_i) = \int_{-\infty}^{\mathbf{x}_i\beta} \phi(t)dt = \Phi(\mathbf{x}_i\beta)$$

$$\text{Logit} : Prob(d_i = 1|\mathbf{x}_i) = \int_{-\infty}^{\mathbf{x}_i\beta} f(t)dt = \frac{e^{\mathbf{x}_i\beta}}{1+e^{\mathbf{x}_i\beta}}$$

With that, it is easy to construct the log-likelihood over the entire sample:

$$ln(L(\beta|\mathbf{d}, \mathbf{x}_i\beta)) = \sum_{i=1}^{N} ln[Prob(d_i = 1|\mathbf{x}_i\beta) \times (d_i) + (1 - Prob(d_i = 1|\mathbf{x}_i\beta)) \times (1 - d_i)]$$

Show:

1. If Logit and $d_i = 1$, person i's contribution to the likelihood function.
2. If Logit and $d_i = 0$, person i's contribution to the likelihood function.
3. Item (2) can be written as $\frac{1}{1+e^{\mathbf{x}_i\beta}}$

Interpreting Parameters

We have focused on marginal effects and elasticities for the models in the class. In an OLS setting, marginal effects are

$$\frac{\partial E(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} \tag{5}$$

For all of the models considered thus far, our estimated parameters are our marginal effects. Here, since the expected value of d_i is:

$$E(d_i|\mathbf{x_i}) = 0 \times (1 - Prob(d_i = 1|\mathbf{x}_i)) + 1 \times Prob(d_i = 1|\mathbf{x}_i)$$

The marginal effect for individual i and regressor K is

$$\frac{\partial E(d_i|\mathbf{x}_{iK})}{\partial \mathbf{x}_{iK}} = f(\mathbf{x}_i\beta)\beta_K$$

or, how the probability of choosing 'Yes' changes when the value of $\mathbf{x}_{iK}$ changes.

Job Choice Example from Mroz

Focus on the decision to be "Working" or "Not Working".

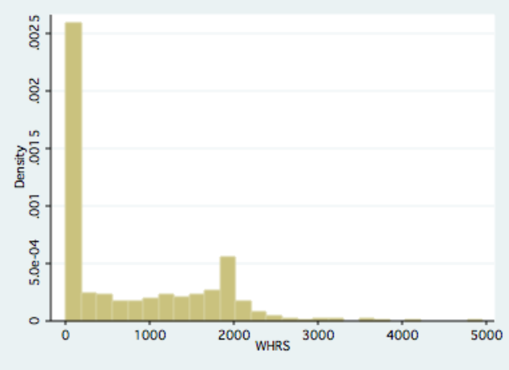

Figure: Actual Working Hours (WHRS)

A Discrete Model of Labor Force Participation

Specify a model to explain the variable *lfp* $= 1$ if the woman worked in 1975, $= 0$ if not. The Data:

LFP	= 1 if woman worked in 1975, else 0,
WHRS	= Wife's hours of work in 1975,
KL6	= Number of children less than 6 years old in household,
K618	= Number of children between ages 6 and 18 in household,
WA	= Wife's age,
WE	= Wife's educational attainment, in years,
WW	= Wife's average hourly earnings, in 1975 dollars,
RPWG	= Wife's wage reported at the time of the 1976 interview (not = 1975 estimated wage),
HHRS	= Husband's hours worked in 1975,
HA	= Husband's age,
HE	= Husband's educational attainment, in years,
HW	= Husband's wage, in 1975 dollars,
FAMINC	= Family income, in 1975 dollars,
WMED	= Wife's mother's educational attainment, in years,
WFED	= Wife's father's educational attainment, in years,
UN	= Unemployment rate in county of residence, in percentage points,
CIT	= Dummy variable = 1 if live in large city (SMSA), else 0,
AX	= Actual years of wife's previous labor market experience

Source. 1976 Panel Study of Income Dynamics, Mroz (1987)

From Mroz: Comparing OLS, Probit, and Logit

The OLS model is sometimes called the Linear Probability Model and can perform fairly well in a wide variety of settings. The problems with the OLS in this case is:

1. The predicted value from an OLS regression ($\hat{\mathbf{d}} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{d}$ is not constrained in the interval [0,1].
2. The estimated marginal effect, $\frac{\partial E(\hat{\mathbf{d}}|\mathbf{x})}{\partial \mathbf{x}} = \mathbf{b}$
3. Errors can't be normally distributed
4. Errors are heteroskedastic

BUT: OLS estimates are unbiased.

From Mroz: Comparing OLS, Probit, and Logit

VARIABLES	(1) OLS	(2) Probit	(3) Logit
kl6	-0.296***	-0.830***	-1.367***
	(0.0367)	(0.110)	(0.189)
k618	-0.0262*	-0.0796**	-0.130**
	(0.0142)	(0.0397)	(0.0654)
faminc	4.14e-06***	1.12e-05***	1.89e-05***
	(1.41e-06)	(3.94e-06)	(6.74e-06)
wa	-0.0153***	-0.0429***	-0.0700***
	(0.00257)	(0.00736)	(0.0122)
Constant	1.228***	2.054***	3.331***
	(0.126)	(0.364)	(0.607)
Observations	753	753	753
R-squared	0.097		

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Comparing OLS, Logit and Probit, cont.

- The models tell a consistent story for signs and significant parameters.
- To compare models, use the approximation
  - Logit-OLS: Scale the logit coefficients by .25: $b^{ols}_{educ} \sim b^{logit}_{educ} \times .25$
  - Probit-OLS: Scale the probit coefficients by .4: $b^{ols}_{educ} \sim b^{probit}_{educ} \times .4$
  - Logit-Probit: Scale the logit coefficient by .6: $b^{probit}_{educ} \sim b^{logit}_{educ} \times .6$

Marginal Effects for Children < 6

# of children	OLS	Probit	Logit
0	-.296	-.308	-.311
1	-.296	-.299	-.298
2	-.296	-.146	-.131
3	-.296	-.036	-.039

Endogeneity

The probit model can be used to test for endogeneity. For example, in the Mroz data we could see if the variable we (wife's education) is endogenous. The steps for testing for H_0: Exogenous Regressor is as before.

- Pick a candidate instrument and test for relevancy
- If relevant, include in regression and then test for exogeneity

The stata command ivprobit will run this regression and test for exogeneity of the regressor.