

EC 6310:

# Advanced Econometric Theory

July 2008

Slides for Lecture on

## Qualitative and Limited Dependent Variable Models

Gary Koop, University of Strathclyde

# 1 Summary

- Readings: Chapter 9 of textbook.
- I will cover tobit, probit, ordered probit and (if time permits) multinomial probit.
- All have general format: conditional on some latent data, model becomes Normal linear regression model.
- Thus, Gibbs sampling with data augmentation can be done.

## 2 Overview

- The Normal linear regression model assumes the distribution of  $y$  given  $X$  to be Normal.
- If data is a qualitative choice or censored, this assumption is not reasonable.
- However, for most models Bayesian inference is straightforward using Gibbs sampling with data augmentation
- We will work with Normal linear regression model with one minor change in notation:  $y^* = (y_1^*, \dots, y_N^*)'$  now denotes the dependent variable:

$$y_i^* = x_i' \beta + \varepsilon_i$$

where  $x_i = (1, x_{i2}, \dots, x_{ik})'$  or, in matrix notation,

$$y^* = X\beta + \varepsilon.$$

- If  $y^*$  were observed, then analysis would proceed as in Chapter 3 or 4. E.g. if we use an independent Normal-Gamma prior for  $\beta$  and  $h$ , then a Gibbs sampler which sequentially draws from  $p(\beta|y^*, h)$  and  $p(h|y^*, \beta)$  could be used to carry out posterior inference.
- All models in this lecture assume that  $y^*$  is unobserved, latent data which is related to the actual, observed data,  $y$  in some manner.
- If we use the independent Normal-Gamma prior for  $\beta$  and  $h$ , then the posterior simulators which result all draw from  $p(\beta|y^*, h)$ ,  $p(h|y^*, \beta)$  and  $p(y^*|y, \beta, h)$  (with natural conjugate prior things are even a bit simpler)

- The only new derivation that is required is  $p(y^*|y, \beta, h)$ .  
That is,  $p(\beta|y^*, h)$  and  $p(h|y^*, \beta)$  are exactly as described in Chapter 4, Section 4.2.
- Accordingly, we will focus on  $p(y^*|y, \beta, h)$ .

### 3 The Tobit Model

- The tobit model is an example of a model which can be used when the data is censored:

$$\begin{aligned} y_i &= y_i^* \text{ if } y_i^* > 0 \\ y_i &= 0 \text{ if } y_i^* \leq 0 \end{aligned} \cdot$$

- Given comments in "Overview", we need only derive  $p(y^*|y, \beta, h)$  and a method for drawing from it in order to complete the specification of a Gibbs sampler for carrying out posterior inference in the tobit model.
- Note first that we have assumed the errors to be independent of one another and, hence, the latent data will exhibit the same property. Thus we can write:

$$p(y^*|y, \beta, h) = \prod_{i=1}^N p(y_i^*|y_i, \beta, h),$$

and focus on  $p(y_i^*|y_i, \beta, h)$ .

- Two cases must be considered:  $y_i > 0$  and  $y_i = 0$ .
- The first of these is simple: if  $y_i > 0$  then we have  $y_i^* = y_i$ . Formally, if  $y_i > 0$ , then the conditional posterior for  $y_i^*$  is a degenerate density with all probability located at the point  $y_i^* = y_i$ .
- The second case can be dealt with by combining the fact that, unconditionally,  $y_i^*$  is Normal with the fact that  $y_i = 0$  implies  $y_i^* \leq 0$ . Thus,  $y_i^*$  has a truncated Normal distribution if  $y_i = 0$ . Formally, we can write  $p(y_i^*|y_i, \beta, h)$  as:

$$y_i^* = y_i \text{ if } y_i > 0$$

$$y_i^* | y_i, \beta, h \sim N(x_i' \beta, h^{-1}) \mathbf{1}(y_i^* < 0) \text{ if } y_i = 0$$

- Empirical illustration provided in textbook.

## 4 The Probit Model

- For the probit model, the relationship between  $y$  and  $y^*$  takes the form:

$$\begin{aligned} y_i &= 1 \text{ if } y_i^* \geq 0 \\ y_i &= 0 \text{ if } y_i^* < 0 \end{aligned} \cdot$$

- Motivation: assume that an individual has to make a choice between two alternatives. Let  $U_{ji}$  be the utility that individual  $i$  (for  $i = 1, \dots, N$ ) associates with choice  $j$  (for  $j = 0, 1$ ).
- The individual makes choice 1 if  $U_{1i} \geq U_{0i}$  and makes choice 0 otherwise.
- Thus, the choice depends on the difference in utilities across the two alternatives and we define this difference as:

$$y_i^* = U_{1i} - U_{0i}.$$

- Probit model assumes difference in utilities follows the Normal linear regression model.
- I.e. individual's utility difference depends on observed characteristics contained in  $x_i$  (e.g. distance to work, salary level, etc.) plus a error which is assumed to be Normally distributed.
- Probit model, and other similar models, are referred to as *random utility models*.
- For reasons discussed above, all we need to derive is  $p(y^*|y, \beta, h)$ .
- Independence across individuals implies that

$$p(y^*|y, \beta, h) = \prod_{i=1}^N p(y_i^*|y_i, \beta, h),$$

and thus, we can focus on  $p(y_i^*|y_i, \beta, h)$ .

- The assumptions of the Normal linear regression model imply that  $p(y_i^*|\beta, h)$  is Normal. Combine this Normality result with information contained in  $y_i$ . If  $y_i = 1$  we obtain a Normal distribution truncated to the left at 0. If  $y_i = 0$  we obtain a Normal distribution truncated to the right at 0.
- To be precise,

$$\begin{aligned} y_i^*|y_i, \beta, h &\sim N(x_i'\beta, h^{-1})\mathbf{1}(y_i^* \geq 0) \text{ if } y_i = 1 \\ y_i^*|y_i, \beta, h &\sim N(x_i'\beta, h^{-1})\mathbf{1}(y_i^* < 0) \text{ if } y_i = 0. \end{aligned} \quad (*)$$

- Posterior inference in the probit model can be carried out using a Gibbs sampler with data augmentation which sequentially draws from (\*) and  $p(\beta, h|y^*)$ .
- Reminder: draws from  $p(\beta, h|y^*)$  can be obtained using our previous results (see Chapters 3 or 4 from textbook).

## 4.1 Features of Interest in Probit Model

- Researchers often interested in presenting choice probabilities which can be derived as follows:

$$\begin{aligned}\Pr(y_i = 1|\beta, h, x_i) &= \Pr(y_i^* \geq 0|\beta, h, x_i) = \\ \Pr(x_i'\beta + \varepsilon_i \geq 0|\beta, h, x_i) &= \Pr(\sqrt{h}\varepsilon_i \geq -\sqrt{h}x_i'\beta|\beta, h, x_i) \\ &= 1 - \Phi(-\sqrt{h}x_i'\beta)\end{aligned}$$

where  $\Phi(\cdot)$  as the c.d.f. of the standard Normal distribution.

- Choice probabilities are functions of the parameters of the model and, hence, their posterior properties can be calculated using Gibbs sampler output.
- This equation illustrates an *identification problem* that was not present in the tobit model. An identification problem is said to occur if multiple values for

the model parameters give rise to the same value for the likelihood function.

- In the probit model,

$$\Pr(x'_i\beta + \varepsilon_i \geq 0|\beta, h) = \Pr(x'_ic\beta + c\varepsilon_i \geq 0|\beta, h)$$

for any positive constant  $c$ . Since  $c\varepsilon_i$  is a  $N(0, c^2h^{-1})$ , the model is still the same probit model, but with different coefficients and error precision.

- The probit model cannot distinguish  $\beta$  and  $h$  separately, but only the product  $\beta\sqrt{h}$ .
- In the probit model, the standard solution to this problem is to set  $h = 1$ .

## 4.2 Empirical Illustration: The Probit Model

- Small artificial data set:

$$y_i^* = .5x_i + \varepsilon_i$$

- The Gibbs sampler with data augmentation described above is used to provide posterior simulator output. We set  $S_0 = 1,000$  and  $S_1 = 10,000$ .

	Posterior Mean	Posterior St. Dev
$\beta_1$	-0.10	0.28
$\beta_2$	0.42	0.48
Probability of Making Choice 1		
Individual with $x_i = 2$	0.73	0.28
Individual with $x_i = 0$	0.46	0.11
Individual with $x_i = -2$	0.27	0.20

- Results near value used true values and can make statements like: "The point estimate indicates that there is a 73% chance that an individual with  $x_i = 2$  will choose alternative 1". Note also that such statements can also be interpreted as predictive choice probabilities for individuals with  $x_i = 2, 0$  or  $-2$ .

## 5 The Ordered Probit Model

- An extension of the probit model to allow for many choices where the choices are ordered (e.g. marketing surveys: consumers asked about product and must choose from alternatives Very Bad, Bad, Indifferent, Good, Very Good).
- As before, we have  $y_i^*$  being latent data with Normal linear regression model:

$$y_i^* = x_i' \beta + \varepsilon_i$$

- Now  $y_i$  can take on values  $\{j = 1, \dots, J\}$  where  $J$  is the number of ordered alternatives and we have:

$$y_i = j \text{ if } \gamma_{j-1} < y_i^* \leq \gamma_j,$$

where  $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_J)'$  is a vector of parameters with  $\gamma_0 \leq \dots \leq \gamma_J$ .

- Impose the identifying restriction  $h = 1$ .
- To aid in interpretation of  $\gamma$ , note that choice probabilities have the form:

$$\begin{aligned} \Pr(y_i = j | \beta, \gamma, x_i) &= \\ \Pr(\gamma_{j-1} < y_i^* \leq \gamma_j | \beta, \gamma, x_i) &= \\ \Pr(\gamma_{j-1} < x_i' \beta + \varepsilon_i \leq \gamma_j | \beta, \gamma, x_i) &= \quad . \\ \Pr(\gamma_{j-1} - x_i' \beta < \varepsilon_i \leq \gamma_j - x_i' \beta | \beta, \gamma, x_i) &= \\ \Phi(\gamma_j - x_i' \beta) - \Phi(\gamma_{j-1} - x_i' \beta) & \end{aligned}$$

- Ordered probit calculates choice probabilities for any individual by taking a Normal distribution and choosing  $\gamma_0, \dots, \gamma_J$  in such a way as to divide up the probability across choices.

- Draw figure of Normal distribution to illustrate identification problem. The standard way of solving this problem is to set  $\gamma_0 = -\infty$ ,  $\gamma_1 = 0$  and  $\gamma_J = \infty$ .
- Bayesian inference in the ordered probit model can be done using a Gibbs sampler with data augmentation which sequentially draws from  $p(\beta|y^*, \gamma)$ ,  $p(\gamma|y^*, y, \beta)$  and  $p(y^*|y, \beta, \gamma)$ .
- As with probit model,  $p(\beta|y_i^*, \gamma)$  will be a Normal density if a Normal or non-informative prior is used for  $\beta$  (remember, we are imposing  $h = 1$ ).
- As with probit model,  $p(y_i^*|y_i, \beta, \gamma)$  is a truncated Normal density:

$$y_i^*|y_i = j, \beta, \gamma \sim N(x_i'\beta, \mathbf{1})\mathbf{1}(\gamma_{j-1} < y_i^* \leq \gamma_j).$$

- The new feature of the ordered probit model is  $p(\gamma|y_i^*, y_i, \beta)$ . We use a flat, improper prior for each of these parameters (i.e.  $p(\gamma_j) \propto c$ ) below, although other priors can be added with minor modifications.
- It proves easier to draw from the components of  $\gamma$  one at a time. Remembering that identification requires  $\gamma_0 = -\infty, \gamma_1 = 0$  and  $\gamma_J = \infty$ , it is convenient to draw from  $p(\gamma_j|y^*, y, \beta, \gamma_{(-j)})$  for  $j = 2, \dots, J - 1$ . The notation  $\gamma_{(-j)}$  denotes  $\gamma$  with  $\gamma_j$  removed (i.e.  $\gamma_{(-j)} = (\gamma_0, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_J)'$ ).
- $p(\gamma_j|y^*, y, \beta, \gamma_{(-j)})$  can be derived by combining a few simple facts. First, we are conditioning on  $\gamma_{(-j)}$  and, thus, we know  $\gamma_j$  must lie in the interval  $[\gamma_{j-1}, \gamma_{j+1}]$ . Second, we are conditioning on both  $y$  and  $y^*$  and, thus, can figure out which values of the latent data correspond to which values for the actual data. Third, the conditioning arguments provide no other information about  $\gamma_j$ . These facts imply a Uniform distribution:

$$\gamma_j | \mathbf{y}^*, \mathbf{y}, \beta, \gamma_{(-j)} \sim U(\bar{\gamma}_{j-1}, \bar{\gamma}_{j+1}),$$

for  $j = 2, \dots, J - 1$ , where

$$\bar{\gamma}_{j-1} = \max \left\{ \max \{y_i^* : y_i = j\}, \gamma_{j-1} \right\},$$

and

$$\bar{\gamma}_{j+1} = \min \left\{ \min \{y_i^* : y_i = j + 1\}, \gamma_{j+1} \right\}.$$

The notation  $\max \{y_i^* : y_i = j\}$  denotes the maximum value of the latent data over all individuals who have chosen alternative  $j$ .  $\min \{y_i^* : y_i = j + 1\}$  is defined in an analogous fashion.

## 6 The Multinomial Probit Model

- *Multinomial probit* model is commonly-used model when several unordered alternatives are involved.
- Assume there are  $J + 1$  alternatives and  $y_i$  can take on values  $\{j = 0, \dots, J\}$ .
- Motivate using random utility model.
- Let  $U_{ji}$  be the utility of individual  $i$  when making choice  $j$ . All information relating to what choices were actually made is provided in utility differences relative to some base alternative (say choice 0) and define latent utility difference variable as:

$$y_{ji}^* = U_{ji} - U_{0i},$$

- As before, assume Normal linear regression model:

$$y_{ji}^* = x'_{ji}\beta_j + \varepsilon_{ji},$$

where  $x_{ji}$  is a  $k_j$ -vector containing explanatory variables which influence the utility associated with choice  $j$  (relative to choice 0),  $\beta_j$  is the corresponding vector of regression coefficients and  $\varepsilon_{ji}$  is the regression error.

- Posterior simulator combines results for the (SUR) model with a method for drawing the latent utility differences (using truncated Normal).
- Put model the form of a SUR model (see Chapter 6, Section 6.6). Thus, we stack all equations into vectors/matrices as for SUR. For instance,  $y_i^* = (y_{1i}^*, \dots, y_{Ji}^*)'$ , etc. to obtain

$$y_i^* = X_i\beta + \varepsilon_i.$$

- Then, if we stack all observations together as for SUR, we obtain

$$y^* = X\beta + \varepsilon.$$

- If make same assumptions about errors as for SUR model, we can use our old SUR results to draw from  $p(\beta|y^*, H)$  (which is Normal) and  $p(H|y^*, \beta)$  (which is Wishart).
- Gibbs sampler completed by figuring out a method for drawing from  $p(y^*|y, \beta, H)$ . It turns out this is truncated Normal.
- The econometrician does not observe  $y_{ji}^*$  directly, but rather observes  $y_i$  where

$$y_i = 0 \text{ if } \max(y_i^*) < 0$$

$$y_i = j \text{ if } \max(y_i^*) = y_{ij}^* \geq 0 \quad ,$$

where  $\max(y_i^*)$  is the maximum of the  $J$ -vector  $y_i^*$ .

- Methods for drawing from  $p(y^*|y, \beta, H)$  can be developed in a similar manner as for the tobit or probit models.
- Independence across individuals implies that

$$p(y^*|y, \beta, H) = \prod_{i=1}^N p(y_i^*|y_i, \beta, H),$$

and, thus, we can focus on  $p(y_i^*|y_i, \beta, H)$ .

- We know  $p(y_i^*|\beta, H)$  is a Normal density. Combining this with the information contained in  $y_i$  yields a truncated Normal density. Thus,

$$y_i^*|y_i, \beta, H \sim N(X_i'\beta, H^{-1})\mathbf{1}(\max(y_i^*) < 0) \text{ if } y_i = 0$$

$$y_i^*|y_i, \beta, H \sim N(X_i'\beta, H^{-1})\mathbf{1}(\max(y_i^*) = y_{ij}^* \geq 0) \text{ if } y_i = j$$

for  $j = 1, \dots, J$ .

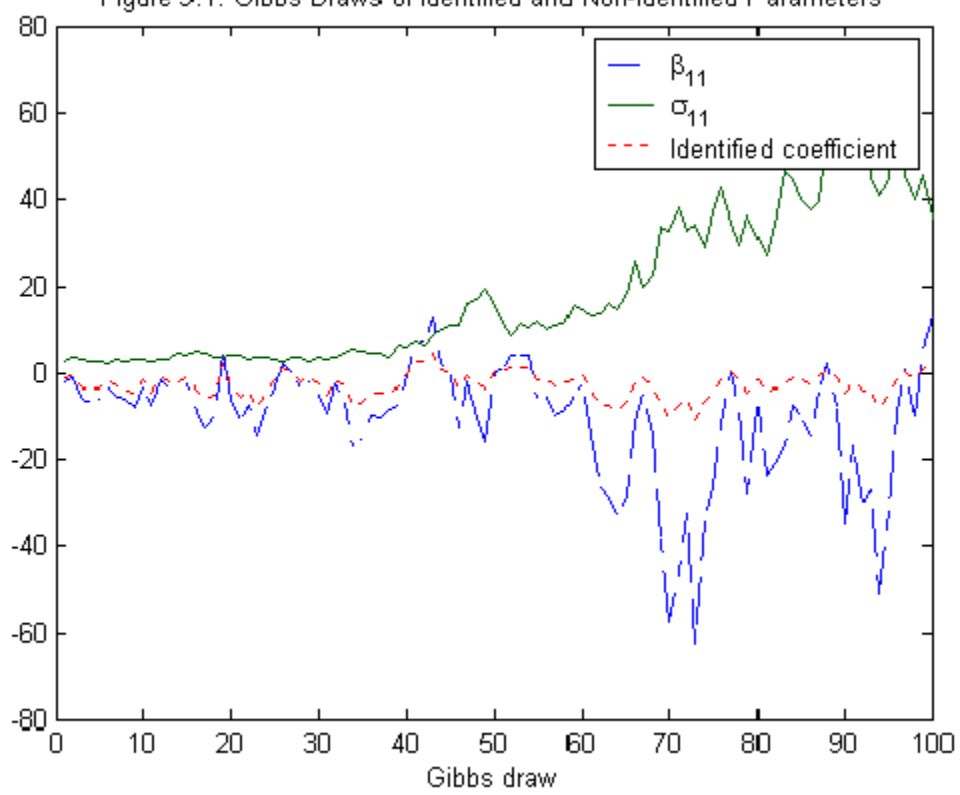
- Thus, Bayesian inference in the multinomial probit model can be carried out using a Gibbs sampler with data augmentation involving the Normal, truncated Normal and Wishart distributions.

## 6.1 Identification in Multinomial Probit Model

- In univariate probit model, we used the identifying restriction  $h = 1$ .
- With the multinomial probit model, imposing the identification restriction is more difficult to do and complicates computation.
- The reason why the multinomial probit model is unidentified is the same as that presented for the probit model (and so we will not repeat motivation).
- If we define the error covariance matrix as  $\Sigma = H^{-1}$  and let  $\sigma_{ij}$  denote the  $ij^{th}$  element of  $\Sigma$ , then the standard method of imposing identification is to set  $\sigma_{11} = 1$ .

- With this restriction imposed  $p(H|y^*, \beta)$  no longer has a Wishart distribution and the results for the SUR model of Chapter 6 can no longer be directly used.
- Several ways of addressing this problem have been proposed in the literature (see textbook pages 224-226 describes one approach).
- However, one simple approach is to ignore the identification issue and do not impose  $\sigma_{11} = 1$ . In this case,  $p(H|y^*, \beta)$  is Wishart and computation is straightforward. Instead of presenting empirical results for  $\beta$ , present results for the identified combination of parameters:  $\frac{\beta}{\sqrt{\sigma_{11}}}$ .
- Figure 9.1 shows what happens from empirical illustration:

Figure 9.1: Gibbs Draws of Identified and Non-identified Parameters



## 6.2 Empirical Illustration: The Multinomial Probit Model

- Details of application given in textbook, pages 226-229)
- We use a data set taken from a marketing application:  $N = 136$  households choosing between four brands of crackers, Sunshine, Keebler, Nabisco and Private label.
- Thus, the data set contains four alternatives. "Private label" is the omitted alternatives (i.e. the latent data,  $y^*$ , contains the difference in utility between the other three brands and Private label).
- For every alternative, we use an intercept and the price of all four brands of crackers in the store at time of purchase as explanatory variables. Thus  $k_1 = k_2 = k_3 = 5$ .

- Posterior inference is based on a multinomial probit model where identifying restriction,  $\sigma_{11}$  is NOT imposed (see previous lecture material) and one where identifying restriction is imposed (see pages 224-226 of textbook)
- Priors are a bit different (which accounts for some difference in results), but overall results with two approaches are quite similar.

Table 9.3: Posterior Results for Multinomial Probit Models				
	Non-identified Model		Identified Model	
	Mean	Standard Deviation	Mean	Standard Dev
Sunshine				
$\beta_{11}$	-3.92	3.17	-2.94	2.96
$\beta_{12}$	-3.10	2.79	-2.69	2.51
$\beta_{13}$	2.79	2.98	1.96	2.21
$\beta_{14}$	0.33	1.81	0.39	1.84
$\beta_{15}$	2.31	2.21	2.36	2.80
Keebler				
$\beta_{21}$	-1.59	5.76	-2.17	4.54
$\beta_{22}$	-2.43	3.52	-2.48	3.33
$\beta_{23}$	-2.74	3.98	-1.56	4.03
$\beta_{24}$	1.21	4.63	0.44	3.36
$\beta_{25}$	3.01	4.85	2.31	4.20
Nabisco				
$\beta_{31}$	-2.59	2.43	-2.28	2.59
$\beta_{32}$	-0.37	1.02	-0.48	1.27
$\beta_{33}$	2.02	2.05	1.73	2.09
$\beta_{34}$	-0.33	0.98	-0.47	1.36
$\beta_{35}$	1.78	1.60	2.21	1.97

## 7 Extensions

- Many extensions of probit models possible: random coefficients probit, probit models for panel data, multinomial multiperiod probit, etc.
- In these, the model can be written as a Normal linear regression model with latent dependent variable and Bayesian inference using a Gibbs sampler with data augmentation using very similar results to those discussed previously.
- The class of *logit models* is a popular competitor to the probit class. These models have the same random utility model motivation as do probit models and the only difference is that they assume a *logistic distribution* for the model errors instead of the Normal assumption of probit models.

- Bayesian computation in logit models is a bit harder (but still not too difficult).
- Textbook offers some further readings on Bayesian analysis of logit models.

## 8 Summary

- We have introduced several models which can be used when the dependent variable is qualitative or censored.
- We have focussed on cases which can be written in terms of a Normal linear model with a latent dependent variable.
- Conditional on this latent data, results from Chapters 3 and 4 for the Normal linear regression model apply.
- Bayesian inference in a wide class of models can be carried out by developing a Gibbs sampler with data augmentation which combines results for the Normal linear regression model (conditional on latent data) with a distribution for the latent data (conditional on the observed data and parameters).