

Homework 1

Robert Ackerman
University of North Carolina

January 22, 2014

14 - 3

Suppose we use an index formulation for a discrete choice model but it is felt that the latent variable is strictly positive. This is accommodated by supposing the latent variable y^* has exponential density with parameter γ , so the density $f(y^*)$ is $f(y^*) = \gamma^{-1} \exp(-y^*/\gamma)$, with $\gamma = \exp(\mathbf{x}'\beta)$. We observe $y = 1$ if $y^* > \mathbf{z}'\alpha$ and $y = 0$ if $y^* \leq \mathbf{z}'\alpha$

(a)

Give the log-likelihood function for the observed data.

$$\begin{aligned} L &= \prod_{i=1}^N f(y_i|x) \\ \ln L &= \sum_{i=1}^N \ln f(y_i|x) \\ &= \sum_{i=1}^N -\mathbf{x}'\beta - (\mathbf{z}'\alpha) / (\exp(\mathbf{x}'\beta)) \end{aligned}$$

(b)

What is the effect of a one-unit change in x_{ji} on $Pr[y_i = 1]$?

$$\frac{\partial}{\partial x_{ji}} P(y_i = 1) = \beta - (\mathbf{z}'\alpha)x_{ji} / \exp(\mathbf{x}'\beta)$$

(c)

Suppose that $y = 1$ if $y^* > \exp(\mathbf{z}'\alpha)$ and $\mathbf{x} = \mathbf{z}$. Do you see any problems in identifying α and/or β ? Explain your answer. *Even with this additional information, we are not going to be able to separate and thus identify both β and α . We can only determine their joint effect.*

14 - 4

Consider the maximum score estimator with objective functions:

$$S_N(\beta) = \sum_{i=1}^N \left\{ y_i \mathbf{1}(\mathbf{x}'_i \beta > 0) + (1 - y_i) \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right\}$$

$$Q_N(\beta) = \sum_{i=1}^N |y_i - \mathbf{1}(\mathbf{x}'_i \beta > 0)|$$

(a)

Show that:

$$S_N(\beta) = \sum_{i=1}^N \left[\mathbf{1}(y_i = 1) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) + \mathbf{1}(y_i = 0) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right]$$

$$\begin{aligned} S_N(\beta) &= \sum_{i=1}^N \left\{ y_i \mathbf{1}(\mathbf{x}'_i \beta > 0) + (1 - y_i) \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right\} \\ &= \sum_{i=1}^N \left\{ \mathbf{1}(\mathbf{x}'_i \beta > 0) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) + (1 - \mathbf{1}(\mathbf{x}'_i \beta > 0)) \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right\} \\ &= \sum_{i=1}^N \left\{ \mathbf{1}(\mathbf{x}'_i \beta > 0) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) + \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right\} \\ &= \sum_{i=1}^N \left[\mathbf{1}(y_i = 1) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) + \mathbf{1}(y_i = 0) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right] \end{aligned}$$

(b)

Show that:

$$Q_N(\beta) = \sum_{i=1}^N \left[\mathbf{1}(y_i = 1) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) + \mathbf{1}(y_i = 0) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) \right]$$

$$\begin{aligned}
Q_N(\beta) &= \sum_{i=1}^N |y_i - \mathbf{1}(\mathbf{x}'_i \beta > 0)| \\
&= \sum_{i=1}^N |\mathbf{1}(y_i = 1)\mathbf{1}(\mathbf{x}'_i \beta > 0) + \mathbf{1}(y_i = 1)\mathbf{1}(\mathbf{x}'_i \beta \leq 0) - \mathbf{1}(\mathbf{x}'_i \beta > 0)| \\
&= \sum_{i=1}^N |\mathbf{1}(y_i = 1)\mathbf{1}(\mathbf{x}'_i \beta > 0) + (1 - \mathbf{1}(y_i = 0))\mathbf{1}(\mathbf{x}'_i \beta \leq 0) - \mathbf{1}(\mathbf{x}'_i \beta > 0)| \\
&= \sum_{i=1}^N |\mathbf{1}(y_i = 1)\mathbf{1}(\mathbf{x}'_i \beta > 0) + \mathbf{1}(\mathbf{x}'_i \beta \leq 0) + (\mathbf{x}'_i \beta \leq 0) - \mathbf{1}(y_i = 0)\mathbf{1}(\mathbf{x}'_i \beta \leq 0) - \mathbf{1}(\mathbf{x}'_i \beta > 0)| \\
&= \sum_{i=1}^N \left[\mathbf{1}(y_i = 1) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) + \mathbf{1}(y_i = 0) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) \right]
\end{aligned}$$

(c)

Using $\mathbf{1}(y_i = 1) = 1 - \mathbf{1}(y_i = 0)$, show that $Q_N(\beta) = N - S_N(\beta)$

$$\begin{aligned}
Q_N(\beta) &= \sum_{i=1}^N \left[\mathbf{1}(y_i = 1) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) + \mathbf{1}(y_i = 0) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) \right] \\
&= \sum_{i=1}^N \left[(1 - \mathbf{1}(y_i = 0)) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) + (1 - \mathbf{1}(y_i = 1)) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) \right] \\
&= N - \sum_{i=1}^N \left[(\mathbf{1}(y_i = 0)) \times \mathbf{1}(\mathbf{x}'_i \beta \leq 0) + \mathbf{1}(y_i = 1) \times \mathbf{1}(\mathbf{x}'_i \beta > 0) \right] \\
&= N - S_N(\beta)
\end{aligned}$$

(d)

Using $\mathbf{1}(\mathbf{x}'_i \beta \leq 0) = 1 - \mathbf{1}(\mathbf{x}'_i \beta > 0)$ show that $S_N(\beta)$ can be rewritten as:

$$\begin{aligned}
S_N(\beta) &= \sum_{i=1}^N (2y_i - 1)\mathbf{1}(\mathbf{x}'_i \beta > 0) + N - \sum_{i=1}^N \\
&= \sum_{i=1}^N \left\{ y_i \mathbf{1}(\mathbf{x}'_i \beta > 0) + (1 - y_i) \mathbf{1}(\mathbf{x}'_i \beta \leq 0) \right\} \\
&= \sum_{i=1}^N \left\{ y_i \mathbf{1}(\mathbf{x}'_i \beta > 0) + (1 - y_i) \mathbf{1}(1 - \mathbf{x}'_i \beta > 0) \right\} \\
&= \sum_{i=1}^N (2y_i - 1)\mathbf{1}(\mathbf{x}'_i \beta > 0) + N - \sum_{i=1}^N
\end{aligned}$$

14 - 5

Use the health expenditure data of Section 16.6. The model is a probit regression of DMED, an indicator variable for positive health expenditures, against just one regressor for simplicity, NDISEASE, the number of chronic diseases.

(a)

Obtain the OLS estimate of the slope parameter.

(See table below for results, and attached for code)

(b)

Obtain the probit estimate of the slope parameter.

(See table below for results, and attached for code)

(c)

Given part (b), obtain the marginal effect of chronic diseases in two ways: averaged over the sample and evaluated at the sample average of NDISEASE.

(See table below for results, and attached for code)

(d)

Obtain the logic estimate of the slope parameter.

(See table below for results, and attached for code)

(e)

Given part (d), obtain the marginal effect of chronic diseases in three ways: averaged over the sample, evaluated at the sample average of NDISEASE, and evaluated at $\Lambda(\mathbf{x}, \beta) = \bar{y}$.

(See table below for results, and attached for code)

(f)

For the logic model calculate the proportionate change in the odds ratio when NDISEASE changes.

14-5 Results			
	<u>OLS</u>	<u>Probit</u>	<u>Logit</u>
Estimate	0.0091	0.0351	0.0615
	(0.0004)	(0.0017)	(0.0030)
P-value	0.000	0.000	0.000
AME	0.0091	0.0102	0.0103
	(0.0005)	(0.0005)	(0.0005)
P-value	0.000	0.000	0.000
MEM	N/A	0.0102	0.0103
	N/A	(0.0005)	(0.0005)
P-value	N/A	0.000	0.000
MER	N/A	N/A	?
	N/A	N/A	(0.0030)
P-value	N/A	N/A	0.000
Log-likelihood	N/A	-10410.18	-10410.89

Note: numbers in parenthesis are standard errors.

14 - 6

Continue the analysis of Exercise 14.5.

(a)

Compare the three binary models on the basis of statistical significance of NDISEASE.

We can't properly interpret significance given that we can only interpret the sign of coefficient, and not magnitude.

(b)

Compare the three binary models on the basis of the estimated marginal effect.

The marginal effects are nearly the same, across all three models.

(c)

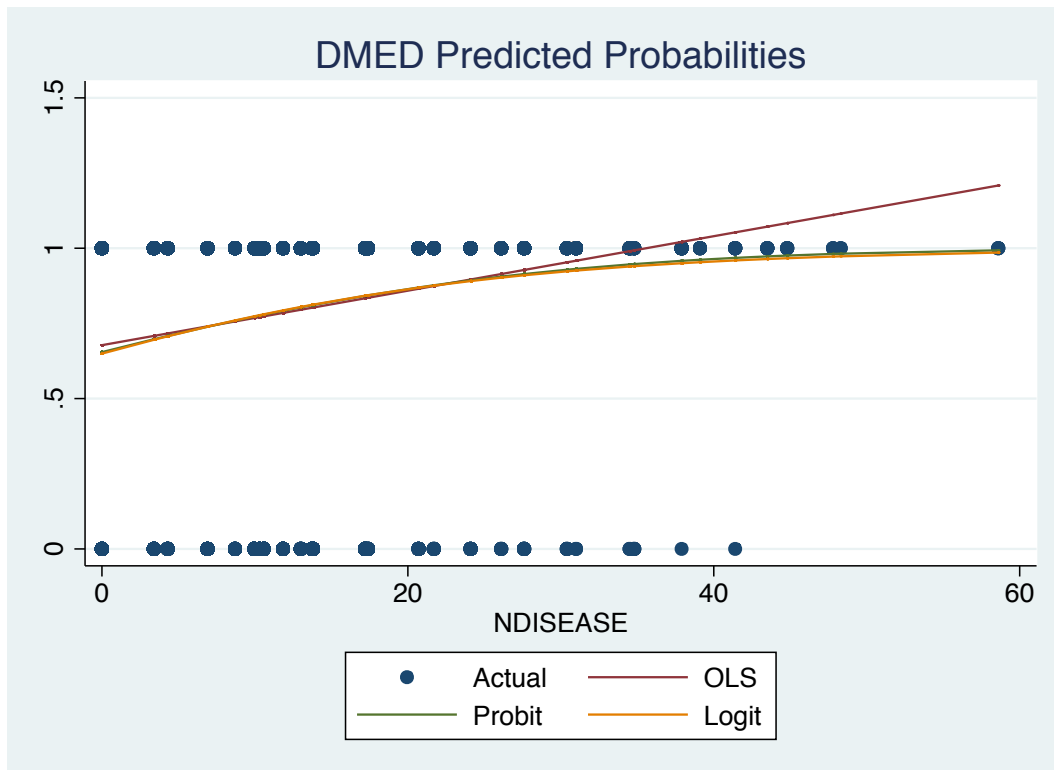
Compare the three binary models on the basis of predicted probabilities

In this regard, the models are a bit different. As the figure below demonstrates, all three predict medical expenditures as an increasing function of the number of chronic diseases. As expected, for large values of NDISEASE the OLS predictions lie outside the interval (0,1). The Probit/Logit results are very similar, and all predicted probabilities fall nicely in (0,1).

(d)

Compare the logic and probit binary models on the basis of the log-likelihood.

As expected, the log-likelihood values for probit/logit are nearly identical.



```
1 clear
2 clear matrix
3 capture cd "/Users/robertackerman/Desktop/Dropbox
4 log using "HW1_Ackerman.log", replace
5
6 use "/Users/robertackerman/Desktop/Dropbox,
7 set more off
8 pause on
9
10 ** Re-name variables (from CT do file) **
11 rename meddol MED
12 rename binexp DMED
13 rename lnmeddol LNMED
14 rename linc LINC
15 rename lfam LFAM
16 rename educdec EDUCDEC
17 rename xage AGE
18 rename female FEMALE
19 rename child CHILD
20 rename fchild FEMCHILD
21 rename black BLACK
22 rename disea NDISEASE
23 rename physlm PHYS LIM
24 rename hlthg HLTHG
25 rename hlthf HLTHF
26 rename hlthp HLTHP
27 rename idp IDP
28 rename logc LC
29 rename lpi LPI
30 rename fmde FMDE
31
32 ** Part 14-5 (a) OLS **
33 reg DMED NDISEASE
34 predict pols, xb
35 margins, dydx(NDISEASE)
36 ** Parts 14-5 (b-c) Probit **
37 probit DMED NDISEASE
38 predict pprobit, pr
39 margins, dydx(NDISEASE)
40 margins, dydx(NDISEASE) atmean
41
42 ** Parts 14-5 (d-f) Logit **
43 logit DMED NDISEASE
44 predict plogit, pr
45 margins, dydx(NDISEASE)
46 margins, dydx(NDISEASE) atmean
47
48 ** Predicted Probabilities Plot **
```

```
49  sort NDISEASE
50  twoway (scatter DMED NDISEASE) (line pols NDISEASE) (line pprobit
NDISEASE) (line plogit NDISEASE)
51  log close
```