Harit Hatawong

6104640112

Assignment 6

The model

In the study of default probability of the loan, determination factors include:

$$Prob(Y=1|X) = f(X_1, X_2, X_3, X_4)$$

Dependent variable $Yi = 1$ if the firm is bad loan, and $= 0$ for good loan.

Independent variables

X_1 is debt coverage ratio.

X_2 is liquidity ratio represented by current assets to current liabilities

X_3 is profitability ratio represented by sales to total assets

X_4 is solidity ratio represented by retained earnings to total assets

From Data `assign6.dta`:

Requirements:

a. Estimate the model assuming that the probability function is (a) cumulative normal probability distribution function and (b) logistic probability distribution function. Interpret your estimated result (overall test, individual test, pseudo R^2, counted R^2).
b. Make comparison of the goodness of fit of the two models.
c. From Probit model, show how to compute Overall LR-test.
d. From Logit model, compute predicted value of index value and predicted probability of being bad loan by using mean value of all *Xs*.
e. Compute marginal effect at mean and at median for Logit model.
f. Compute marginal effect at the value of X_1=0.5, X_2=1, X_3=0.5, X_4=0 for the Probit model.
g. Determine counted R^2 using the threshold of predicted value = 0.5 for Logit models.
h. Determine counted R^2 using the threshold of predicted value = 0.7 for Logit models.

a. Estimate the model assuming that the probability function is (a) cumulative normal probability distribution function and (b) logistic probability distribution function. Interpret your estimated result (overall test, individual test, pseudo R^2, counted R^2).

a.) normal distribution (Probit model)

$$P(Y=1|X) = \Phi(x_1, x_2, x_3, x_4)$$

```
2 . probit y x1 x2 x3 x4

Iteration 0:   log likelihood = -248.43455
Iteration 1:   log likelihood = -150.03919
Iteration 2:   log likelihood = -147.48531
Iteration 3:   log likelihood = -147.46882
Iteration 4:   log likelihood = -147.46881

Probit regression                               Number of obs     =        400
                                                LR chi2(4)        =     201.93
                                                Prob > chi2       =     0.0000
Log likelihood = -147.46881                     Pseudo R2         =     0.4064
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.3590739	.0371539	9.66	0.000	.2862536	.4318941
x2	-.8525746	.144481	-5.90	0.000	-1.135752	-.569397
x3	-.5735764	.2202882	-2.60	0.009	-1.005333	-.1418195
x4	-1.248569	.226762	-5.51	0.000	-1.693014	-.8041238
_cons	1.45664	.2037279	7.15	0.000	1.057341	1.85594

```
3 . fitstat

Measures of Fit for probit of y

Log-Lik Intercept Only:        -248.435   Log-Lik Full Model:          -147.469
D(395):                         294.938   LR(4):                        201.931
                                          Prob > LR:                      0.000
McFadden's R2:                    0.406   McFadden's Adj R2:              0.386
Maximum Likelihood R2:            0.396   Cragg & Uhler's R2:             0.557
McKelvey and Zavoina's R2:        0.640   Efron's R2:                     0.446
Variance of y*:                   2.775   Variance of error:              1.000
Count R2:                         0.818   Adj Count R2:                   0.416
AIC:                              0.762   AIC*n:                        304.938
BIC:                          -2071.691   BIC':                        -177.966
```

overall test: $p < \alpha = 0.05$

jointly significant in overall test

individual test: (z test)

significant at $\alpha = 5\%$

Pseudo R^2 = 0.4064

Count R^2 = 0.818 > 0.5

However adj. Count R^2 < 0.5

so, the model is not accurate predict the y b.) logistic distribution (logit model

$\hat{p}_i = Pr(Y=1|X) = \Lambda(x_1, x_2, x_3, x_4)$

```
7 . logit y x1 x2 x3 x4

Iteration 0:   log likelihood = -248.43455
Iteration 1:   log likelihood = -154.06753
Iteration 2:   log likelihood = -148.00091
Iteration 3:   log likelihood = -147.90887
Iteration 4:   log likelihood = -147.90869
Iteration 5:   log likelihood = -147.90869

Logistic regression                             Number of obs     =        400
                                                LR chi2(4)        =     201.05
                                                Prob > chi2       =     0.0000
Log likelihood = -147.90869                     Pseudo R2         =     0.4046
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.6299401	.0708979	8.89	0.000	.4909828	.7688974
x2	-1.488248	.2597744	-5.73	0.000	-1.997396	-.9790992
x3	-.9562902	.3882611	-2.46	0.014	-1.717268	-.1953124
x4	-2.155321	.4055058	-5.32	0.000	-2.950097	-1.360544
_cons	2.5165	.3714373	6.78	0.000	1.788496	3.244503

```
8 . fitstat

Measures of Fit for logit of y

Log-Lik Intercept Only:      -248.435   Log-Lik Full Model:         -147.909
D(395):                       295.817   LR(4):                       201.052
                                        Prob > LR:                     0.000
McFadden's R2:                  0.405   McFadden's Adj R2:             0.385
Maximum Likelihood R2:          0.395   Cragg & Uhler's R2:            0.555
McKelvey and Zavoina's R2:      0.622   Efron's R2:                    0.445
Variance of y*:                 8.707   Variance of error:             3.290
Count R2:                       0.818   Adj Count R2:                  0.416
```

overall test: $p < \alpha = 0.05$

jointly significant in overall test

individual test (z test)

significant at $\alpha = 5\%$

Pseudo R^2 = 0.4046

Count R^2 = 0.818 > 0.5

However adj. Count R^2 < 0.5

so, the model is not accurate predict the y

b. Make comparison of the goodness of fit of the two models.

From both models pseudo R^2, counted R^2 and log likelihood value. Both models are not significantly different in term of goodness of fit.

c. From Probit model, show how to compute Overall LR-test.

Overall test $\quad H_0: \beta_2 = \beta_3 = \ldots \beta_k = 0$

```
2 . probit y x1 x2 x3 x4

Iteration 0:   log likelihood = -248.43455
Iteration 1:   log likelihood = -150.03919
Iteration 2:   log likelihood = -147.48531
Iteration 3:   log likelihood = -147.46882
Iteration 4:   log likelihood = -147.46881

Probit regression                               Number of obs     =        400
                                                LR chi2(4)        =     201.93
                                                Prob > chi2       =     0.0000
Log likelihood = -147.46881                     Pseudo R2         =     0.4064

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .3590739   .0371539     9.66   0.000     .2862536    .4318941
          x2 |  -.8525746    .144481    -5.90   0.000    -1.135752    -.569397
          x3 |  -.5735764   .2202882    -2.60   0.009    -1.005333   -.1418195
          x4 |  -1.248569    .226762    -5.51   0.000    -1.693014   -.8041238
       _cons |    1.45664   .2037279     7.15   0.000     1.057341     1.85594
------------------------------------------------------------------------------

4 . probit y, nolog

Probit regression                               Number of obs     =        400
                                                LR chi2(0)        =       0.00
                                                Prob > chi2       =          .
Log likelihood = -248.43455                     Pseudo R2         =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4887764   .0654634     7.47   0.000     .3604706    .6170822
------------------------------------------------------------------------------
```

test stat $= 2(\ln L_{UR} - \ln L_R) \sim \chi^2_{(k-1)}$

$= 2(-147.469 + 248.435)$

$= 201.93$

d. From Logit model, compute predicted value of index value and predicted probability of being bad loan by using mean value of all *Xs*.

```
9 . mfx, predict(xb)

Marginal effects after logit
      y  = Linear prediction (log odds) (predict, xb)
         =     1.32418
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
x1	.6299401	.0709	8.89	0.000	.490983	.768897	.454973
x2	-1.488248	.25977	-5.73	0.000	-1.9974	-.979099	.809344
x3	-.9562902	.38826	-2.46	0.014	-1.71727	-.195312	.556712
x4	-2.155321	.40551	-5.32	0.000	-2.9501	-1.36054	-.119684

```
10 . mfx

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .7898763
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
x1	.1045522	.01146	9.12	0.000	.082083	.127022	.454973
x2	-.247007	.04388	-5.63	0.000	-.333011	-.161003	.809344
x3	-.1587171	.06397	-2.48	0.013	-.2841	-.033334	.556712
x4	-.3577223	.06679	-5.36	0.000	-.488633	-.226812	-.119684

e. Compute marginal effect at mean and at median for Logit model.

```
10 . mfx

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .7898763
```

at mean

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.1045522	.01146	9.12	0.000	.082083	.127022	.454973
x2	-.247007	.04388	-5.63	0.000	-.333011	-.161003	.809344
x3	-.1587171	.06397	-2.48	0.013	-.2841	-.033334	.556712
x4	-.3577223	.06679	-5.36	0.000	-.488633	-.226812	-.119684

```
11 . mfx, at(median)

Marginal effects after logit
      y  = Pr(y) (predict)
         = .84127022
```

at median

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.0841188	.00961	8.76	0.000	.065292	.102946	.655749
x2	-.1987326	.03349	-5.93	0.000	-.264373	-.133093	.692745
x3	-.1276979	.04944	-2.58	0.010	-.224597	-.030799	.488768
x4	-.28781	.05616	-5.12	0.000	-.397881	-.177739	-.109732

f. Compute marginal effect at the value of X_1=0.5, X_2=1, X_3=0.5, X_4=0 for the Probit model.

```
6 . mfx, at(0.5 1 0.5 0)

Marginal effects after probit
      y  = Pr(y) (predict)
         =      .69034
```

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.1266183	.01307	9.69	0.000	.101001	.152235	.5
x2	-.3006389	.05452	-5.51	0.000	-.4075	-.193778	1
x3	-.2022572	.07644	-2.65	0.008	-.352085	-.05243	.5
x4	-.4402764	.08419	-5.23	0.000	-.605293	-.27526	0

g. Determine counted R^2 using the threshold of predicted value = 0.5 for Logit models.

```
12 . predict pr
   (option pr assumed; Pr(y))

13 . g yhat1=0 if pr<=0.5
   (300 missing values generated)

14 . replace yhat1=1 if pr>0.5
   (300 real changes made)

15 . tabulate y yhat1
```

y	yhat1 0	yhat1 1	Total
0	76	49	125
1	24	251	275
Total	100	300	400

Counted $R^2 = \frac{76+251}{400}$

$= 0.8175 \approx 0.82$ #

h. Determine counted R^2 using the threshold of predicted value = 0.7 for Logit models.

```
16 . g yhat2=0 if pr<=0.7
```

2/3/2564 BE 20:48

it Hatawong

```
   (241 missing values generated)

17 . replace yhat2=1 if pr>0.7
   (241 real changes made)

18 . tabulate y yhat2
```

y	yhat2 0	yhat2 1	Total
0	101	24	125
1	58	217	275
Total	159	241	400

$$\text{counted } R^2 = \frac{101 + 217}{400}$$

$$= 0.795 \#$$