Thanakrit 6204640087

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.

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).

```
.
.          *A*
. *Cumulative Normal Probability Distribution*
. probit y x1 x2 x3 x4, nolog

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
------------------------------------------------------------------------------

. estimates store probit

. 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
```

The estimated result above is when we assume the probability distribution to be a cumulative normal. Then, looking at the overall test - LR χ^2 test, it can be concluded that all independent variables are statistically jointly significant at any levels (P-value = 0.00). The individual test is the z-test because we assume asymptotic normal assumption. Each independent variable is significant at any levels.

Pseudo R^2 = 0.406 and Counted R^2 = 0.818, these are used for comparison (not for interpretation).

```
. *Logistic Probability Distribution*
. logit y x1 x2 x3 x4, nolog

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
------------------------------------------------------------------------------

. estimates store logit

. 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
AIC:                             0.765   AIC*n:                         305.817
BIC:                         -2070.811   BIC':                         -177.086
```

Instead, we assume probability distribution function to be a logistic. The estimated result changes. By LR - χ^2 test, it also indicates that all regressors are significant. According to the individual test, all regressors are significant individually at any levels except x_3, it will be insignificant if we test at 1% significance level.

Pseudo R^2 = 0.405 , Counted R^2 = 0.818

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

Using Pseudo R^2 and Counted R^2, we should use the probit model (assuming cumulative normal probability distribution) because of a little higher Pseudo - R^2 (which represents goodness of fit of the model " .

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

Firstly, to perform the overall LR-test,

LR-test = $2(\log L_{UR} - \log L_R) \sim \chi^2_{k-1}$ where $\log L_{UR}$ = log likelihood value of unrestricted model.

$\log L_R$ = ——— restricted ———.

```
.              *C*
. probit y x1 x2 x3 x4, nolog
```

=> Unrestricted

```
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

```
. probit y, nolog
```

=> Restricted

```
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

So, LR-test = $2(-147.47 - (-248.43)) = \boxed{201.92 \sim \chi^2}$

Then, use the χ^2 table with degree of freedom $k-1 = 4-1 = 3$, P-value = 0.000.

Thus, it can be concluded that, using Probit model, all independent variables are significant.

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

```
.         *D*
. estimates restore logit
(results logit are active now)

. mfx, predict(xb)

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

$= \hat{I}$

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

```
. mfx

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

$= \hat{P}$

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

From the index function, $\hat{I} = X\hat{\beta} = 1.32$. We can plug in $\hat{I}$ into the probabilitity function $\hat{P}(Y=1 | X) = \frac{1}{1+e^{-X\hat{\beta}}} = \frac{1}{1+e^{-\hat{I}}} = \frac{1}{1+e^{-1.32}} = 0.79$.

Hence, using mean values of all X_s, the probability of being bad loan is predicted to be 0.79 or 79%.

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

```
.         *E*
. mfx

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

Marginal Effect 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

```
. mfx, at(median)

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

Marginal Effect 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

The interpretation is, for example,

" The predicted probability of being bad loan ($\hat{P}(Y=1|x)$) will increase by 0.1045 (10.45%) if x_1 changes by 1 unit from its mean (from 0.45 to 1.45), holding other independent variables at their means."

or

" The predicted probability of being bad loan ($\hat{P}(Y=1|x)$) will increase by 0.0841 (8.41%) if x_1 changes by 1 unit from its median (from 0.65 to 1.65), holding other independent variables at their medians."

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.

```
.         *F*
. estimates restore probit
(results probit are active now)

. 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	-.193777	1
x3	-.2022572	.07644	-2.65	0.008	-.352085	-.05243	.5
x4	-.4402764	.08419	-5.23	0.000	-.605293	-.275259	0

The interpretation is, for example,

" The predicted probability of being bad loan ($\hat{P}(Y=1|x)$) will increase by 0.1266 (12.66%) if x_1 changes by 1 unit from 0.5 to 1.5, holding x_2, x_3, x_4 at 1, 0.5, and 0 respectively "

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

```
.         *G*
. estimates restore logit
(results logit are active now)

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

. gen yhatg = 1 if pr>0.5
(100 missing values generated)

. replace yhatg = 0 if pr<=0.5
(100 real changes made)
```

```
. tabulate y yhatg
```

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

Using Logit Model, I have generated $\hat{Y}=1$ if $\hat{P}>0.5$ and $\hat{Y}=0$ if $\hat{P}\leq 0.5$. Then, tabulate.

Overall Counted $R^2 = \dfrac{\text{Total No. of correct predictions}}{\text{No. of observation}} = \dfrac{76+251}{400} = 81.75\%$

$Y=0$ Counted $R^2 = \dfrac{76}{125} = 60.8\%$

$Y=1$ Counted $R^2 = \dfrac{251}{275} = 91.27\%$

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

```
.
.         *H*
. gen yhatf = 1 if pr>0.7
(159 missing values generated)

. replace yhatf = 0 if pr<=0.7
(159 real changes made)

. tabulate y yhatf
```

y	yhatf 0	1	Total
0	$\boxed{101}$	24	125
1	58	$\boxed{217}$	275
Total	159	241	400

Using Logit Model, I have generated $\hat{Y} = 1$ if $\hat{P} > 0.7$ and $\hat{Y} = 0$ if $\hat{P} \leq 0.7$

Then, tabulate.

$$\text{Overall Counted } R^2 = \frac{\text{Total No. of correct predictions}}{\text{No. of observation}} = \frac{101+217}{400} = 79.5\ \%$$

$$Y=0 \quad \text{Counted } R^2 = \frac{101}{125} = 80.8\ \%$$

$$Y=1 \quad \text{Counted } R^2 = \frac{217}{275} = 78.91\ \%$$

```
                                                  ___  ____  ____  ____  ____(R)
                                                 /__    /   ____/   /   ____/
                                                ___/   /   /___/   /   /___/
                                                  Statistics/Data Analysis
```

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     MP - Parallel Edition            College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Single-user 8-core Stata perpetual license:
       Serial number:  10699393
         Licensed to:  T


Notes:
      1.  Unicode is supported; see help unicode advice.
      2.  More than 2 billion observations are allowed; see help obs advice.
      3.  Maximum number of variables is set to 5000; see help set maxvar.

. use "C:\Users\Chaiyapong M\Desktop\BE\EE426\Assignment 6\assign6.dta", clear

. do "C:\Users\Chaiyapong M\Desktop\BE\EE426\Assignment 6\Assignment 6.do"

. set more off

. 
. *Assignment6 Logit Probit Models*
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
x1	400	.4549731	3.100887	-8.155268	9.74565
x2	400	.8093444	.6100628	.0004192	2.852366
x3	400	.5567118	.3871706	.0016088	1.874493
x4	400	-.119684	.387141	-1.213392	.9788648
x5	400	3.165124	2.190214	.0013003	13.1798
y	400	.6875	.4640929	0	1

```
. 
.         *A*
. *Cumulative Normal Probability Distribution*
. probit y x1 x2 x3 x4, nolog

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

```
. estimates store probit

. 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

```
. *Logistic Probability Distribution*
. logit y x1 x2 x3 x4, nolog
```

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

```
. estimates store logit
```

```
. 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
AIC:                            0.765    AIC*n:                         305.817
BIC:                        -2070.811    BIC':                         -177.086

.
.         *C*
. probit y x1 x2 x3 x4, nolog

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

```
. 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

```
.
.         *D*
. estimates restore logit
(results logit are active now)

. 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

```
. 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*
. 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

```
. mfx, at(median)

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

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*
. estimates restore probit
(results probit are active now)

. 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	-.193777	1
x3	-.2022572	.07644	-2.65	0.008	-.352085	-.05243	.5
x4	-.4402764	.08419	-5.23	0.000	-.605293	-.275259	0

```
.
.         *G*
. estimates restore logit
(results logit are active now)

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

. gen yhatg = 1 if pr>0.5
(100 missing values generated)

. replace yhatg = 0 if pr<=0.5
(100 real changes made)
```

```
. tabulate y yhatg

           |        yhatg
         y |         0          1 |     Total
-----------+----------------------+----------
         0 |        76         49 |       125 
         1 |        24        251 |       275 
-----------+----------------------+----------
     Total |       100        300 |       400 


. 
.         *H*
. gen yhatf = 1 if pr>0.7
(159 missing values generated)

. replace yhatf = 0 if pr<=0.7
(159 real changes made)

. tabulate y yhatf

           |        yhatf
         y |         0          1 |     Total
-----------+----------------------+----------
         0 |       101         24 |       125 
         1 |        58        217 |       275 
-----------+----------------------+----------
     Total |       159        241 |       400 


. 
. 
. 
. 
end of do-file

. 
```