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.

d) a) Overall test (Probit) => LR-chisquare test = 201.93 $\sim \chi^2_4$

Pvalue = 0

It show that all beta are statistically significant

Individuals tests under asymptotic normal ($n \to \infty$)

P-value[z-test] = 0 for x_1, x_2, x_4, constant => all are statistically significant

P-value = 0.009 < 5% confidence level = it also statistically significant

Pseudo R^2 = 0.4064 -> no-interpret only for comparsion

Counted R^2 = 0.8175 -> The model could predict the correct value of y by 81.75%

b) Overall test (Logit) => LR-chisquare test = 201.05 $\sim \chi^2_4$

Pvalue = 0

It show that all beta are statistically significant

Individuals tests under asymptotic normal ($n \to \infty$)

P-value[z-test] = 0 for x_1, x_2, x_4, constant => all are statistically significant

P-value = 0.014 < 0.05 it is statistically significant.

Pseudo R^2 = 0.4064 -> no-interpret only for comparsion

Counted R^2 = 0.8175 -> The model could predict the correct value of y by 81.75%

```
      name:  <unnamed>
       log:  C:\Users\Pongpanot\Downloads\ASS6_Q1.smcl
  log type:  smcl
 opened on:   3 Mar 2021, 23:56:28

1 . use "C:\Users\Pongpanot\Downloads\assign6.dta", clear

2 . 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
------------------------------------------------------------------------------

3 . estat clas

Probit model for y

              -------- True --------
Classified |         D            ~D |      Total
-----------+--------------------------+-----------
         + |       251            49 |        300
         - |        24            76 |        100
-----------+--------------------------+-----------
     Total |       275           125 |        400

Classified + if predicted Pr(D) >= .5
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   91.27%
Specificity                     Pr( -|~D)   60.80%
Positive predictive value       Pr( D| +)   83.67%
Negative predictive value       Pr(~D| -)   76.00%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   39.20%
False - rate for true D         Pr( -| D)    8.73%
False + rate for classified +   Pr(~D| +)   16.33%
False - rate for classified -   Pr( D| -)   24.00%
--------------------------------------------------
Correctly classified                        81.75%
--------------------------------------------------

4 . 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

5 . estat clas

Logistic model for y

	True		
Classified	D	~D	Total
+	251	49	300
-	24	76	100
Total	275	125	400

Classified + if predicted Pr(D) >= .5
True D defined as y != 0

Sensitivity	Pr(+| D)	91.27%
Specificity	Pr(-|~D)	60.80%
Positive predictive value	Pr(D| +)	83.67%
Negative predictive value	Pr(~D| -)	76.00%
False + rate for true ~D	Pr(+|~D)	39.20%
False - rate for true D	Pr(-| D)	8.73%
False + rate for classified +	Pr(~D| +)	16.33%
False - rate for classified -	Pr(D| -)	24.00%
Correctly classified		81.75%

6 . 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

```
7 . mfx

Marginal effects after probit
      y  = Pr(y) (predict)
         =  .77640251
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.1073091	.01086	9.88	0.000	.086023	.128595	.454973
x2	-.2547917	.04347	-5.86	0.000	-.339992	-.169592	.809344
x3	-.1714132	.06552	-2.62	0.009	-.299828	-.042998	.556712
x4	-.3731346	.06687	-5.58	0.000	-.504199	-.24207	-.119684

```
8 . 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

```
9 . 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

```
10 . 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

11 . 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

12 . 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

13 . 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

```
14 . 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
------------------------------------------------------------------------------

15 . 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
------------------------------------------------------------------------------

16 . 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
------------------------------------------------------------------------------

17 . 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

18 . 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

19 . 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
Pseudo R2	=	0.4046

Log likelihood = -147.90869

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

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

21 . g yhat=0 if pr<=0.5
(300 missing values generated)

22 . replace yhat=1 if pr>0.5
(300 real changes made)

```
23 . tabulate y yhat
```

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

```
24 . estat clas

Logistic model for y
```

Classified	True D	~D	Total
+	**251**	**49**	**300**
-	**24**	**76**	**100**
Total	**275**	**125**	**400**

```
Classified + if predicted Pr(D) >= .5
True D defined as y != 0
```

Sensitivity	Pr(+| D)	**91.27%**
Specificity	Pr(-|~D)	**60.80%**
Positive predictive value	Pr(D| +)	**83.67%**
Negative predictive value	Pr(~D| -)	**76.00%**
False + rate for true ~D	Pr(+|~D)	**39.20%**
False - rate for true D	Pr(-| D)	**8.73%**
False + rate for classified +	Pr(~D| +)	**16.33%**
False - rate for classified -	Pr(D| -)	**24.00%**
Correctly classified		**81.75%**

```
25 . 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

```
26 . clear

27 . use "C:\Users\Pongpanot\Downloads\assign6.dta", clear

28 . 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
------------------------------------------------------------------------------

29 .
30 .
31 . predict pr
(option pr assumed; Pr(y))

32 . g yhat=0 if pr<=0.7
(241 missing values generated)

33 . replace yhat=1 ifpr>0.7
invalid 'ifpr'
r(198);

34 . replace yhat=1 if pr>0.7
(241 real changes made)

35 . tabulate y yhat

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

36 . estat clas

Logistic model for y

              -------- True --------
Classified |         D            ~D |     Total
-----------+----------------------+----------
         + |       251            49 |       300
         - |        24            76 |       100
-----------+----------------------+----------
     Total |       275           125 |       400
```

```
Classified + if predicted Pr(D) >= .5
True D defined as y != 0
```

Sensitivity	Pr(+| D)	**91.27%**
Specificity	Pr(-|~D)	**60.80%**
Positive predictive value	Pr(D| +)	**83.67%**
Negative predictive value	Pr(~D| -)	**76.00%**
False + rate for true ~D	Pr(+|~D)	**39.20%**
False - rate for true D	Pr(-| D)	**8.73%**
False + rate for classified +	Pr(~D| +)	**16.33%**
False - rate for classified -	Pr(D| -)	**24.00%**
Correctly classified		**81.75%**

```
37 . log close
      name:  <unnamed>
       log:  C:\Users\Pongpanot\Downloads\ASS6_Q1.smcl
  log type:  smcl
 closed on:   4 Mar 2021, 01:01:56
```