Assignment 3
Solution

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

Requirements:

1. 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).
2. Make comparison of the goodness of fit of the two models.
3. From Probit model, show how to compute Overall LR-test.
4. From Logit model, compute predicted value of index value and predicted probability of being bad loan by using mean value of all *Xs*.
5. Compute marginal effect at mean and at median for Logit model.
6. 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.
7. Determine counted R^2 using the threshold of predicted value = 0.5 for Logit models.
8. Determine counted R^2 using the threshold of predicted value = 0.7 for Logit models.

1 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).

```
. 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
------------------------------------------------------------------------------

. 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
ML (Cox-Snell) R2:               0.396   Cragg-Uhler(Nagelkerke) R2:      0.557
McKelvey & 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
BIC used by Stata:             324.895   AIC used by Stata:             304.938

. 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
------------------------------------------------------------------------------

. 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
ML (Cox-Snell) R2:               0.395   Cragg-Uhler(Nagelkerke) R2:      0.555
McKelvey & 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
BIC used by Stata:             325.775   AIC used by Stata:             305.817
```

2 Make comparison of the goodness of fit of the two models.
According to the above results, it seems that probit model provide a little bit better fitting the data compare to logit model since McFadden's R2 from probit is a little bit higher than from logit `(0.406>0.405)`.

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

From the above result LR-test = `2 ((-147.469) - (-248.435)) = 201.93`

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

```
. 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
------------------------------------------------------------------------------
```

From the estimated result,

$$\hat{I} = 2.52 + 0.62(0.45) - 1.49(0.81) - 0.96(0.56) - 2.16(-0.12) = 1.32$$

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

```
. 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 -.133092   .692745
      x3 |  -.1276979      .04944   -2.58   0.010  -.224597 -.030799   .488768
      x4 |    -.28781      .05616   -5.12   0.000  -.397881 -.177739  -.109732
------------------------------------------------------------------------------
```

6 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.

```
. 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
------------------------------------------------------------------------------
```

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

```
. 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%
--------------------------------------------------
```

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

```
. estat clas, cut(0.7)

Logistic model for y

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       217            24  |        241
     -     |        58           101  |        159
-----------+--------------------------+-----------
   Total   |       275           125  |        400

Classified + if predicted Pr(D) >= .7
True D defined as y != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   78.91%
Specificity                     Pr( -|~D)   80.80%
Positive predictive value       Pr( D| +)   90.04%
Negative predictive value       Pr(~D| -)   63.52%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   19.20%
False - rate for true D         Pr( -| D)   21.09%
False + rate for classified +   Pr(~D| +)    9.96%
False - rate for classified -   Pr( D| -)   36.48%
--------------------------------------------------
Correctly classified                        79.50%
--------------------------------------------------
```