We want to investigate whether the race of a mortgage applicant affects their chances of the application being approved. Mortgage application acceptance is a binary dependent variable.

First we generate some variables:

* deny is a binary variable taking the value 1 if a mortgage application is rejected and 0 if it is not rejected
* pi\_rat shows the debt to income ratio (the banks’ calculation of housing expense/income) divided by 100
* black is a dummy variable taking the value of 1 if the applicant is black.

. set more 1;

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

. \* Read In Data;

. \* (Note: Change path name so that it is appropriate for your computer);

. use hmda\_sw.dta;

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

. gen deny = (s7==3);

. gen pi\_rat = s46/100;

. gen black = (s13==3);

We can generate tables showing the probability of an application being rejected for black and other applicants.

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

. \* Results on Page 1;

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

. sort black;

. summarize deny if (black==1);

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

deny | 339 .2831858 .4512119 0 1

. summarize deny if (black==0);

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

deny | 2041 .0926017 .2899445 0 1

We create some control variables and summarise them

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

> \*\*\*\* Table 11.1 ;

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

. gen hse\_inc = s45/100;

. gen loan\_val = s6/s50;

. gen ccred = s43;

. gen mcred = s42;

. gen pubrec = (s44>0);

. gen denpmi = (s53==1);

. gen selfemp = (s27a==1);

. gen married = (s23a=="M");

. gen single = (married==0);

. gen hischl = (school>=12);

. gen probunmp = uria;

. gen condo = (s51 == 1);

. sum pi\_rat hse\_inc loan\_val ccred mcred pubrec denpmi selfemp

> single hischl probunmp condo black deny;

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

pi\_rat | 2380 .3308136 .1072573 0 3

hse\_inc | 2380 .2553461 .0966556 0 3

loan\_val | 2380 .7377759 .178751 .02 1.95

ccred | 2380 2.116387 1.666721 1 6

mcred | 2380 1.721008 .5372816 1 4

-------------+--------------------------------------------------------

pubrec | 2380 .0735294 .2610584 0 1

denpmi | 2380 .0201681 .1406045 0 1

selfemp | 2380 .1163866 .3207553 0 1

single | 2380 .3932773 .4885802 0 1

hischl | 2380 .9836134 .1269835 0 1

-------------+--------------------------------------------------------

probunmp | 2380 3.774496 2.027062 1.8 10.6

condo | 2380 .2882353 .4530364 0 1

black | 2380 .142437 .3495712 0 1

deny | 2380 .1197479 .3247347 0 1

We also create a list of categorical variables

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

> \*\*\*\* Table 11.2 ;

. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

. gen ltv\_med = (loan\_val>=0.80)\*(loan\_val<=.95);

. gen ltv\_high = (loan\_val>0.95);

. gen blk\_pi = black\*pi\_rat;

. gen blk\_hse = black\*hse\_inc;

. gen ccred3 = (ccred==3);

. gen ccred4 = (ccred==4);

. gen ccred5 = (ccred==5);

. gen ccred6 = (ccred==6);

. gen mcred3 = (mcred==3);

. gen mcred4 = (mcred==4);

First we run a linear probability model

. \*\* col(1);

. quietly regress deny black pi\_rat hse\_inc ltv\_med ltv\_high ccred mcred pubrec denpmi

> selfemp, r;

. eststo LPM;

Then we run a logit model. We can compute the predicted probability for each value of black at the means of all other variables using the –margins- command

. \*\* col(2) - Logit with robust standard errors;

. logit deny i.black `controls', r;

Iteration 0: log pseudolikelihood = -872.0853

Iteration 1: log pseudolikelihood = -672.05096

Iteration 2: log pseudolikelihood = -656.94676

Iteration 3: log pseudolikelihood = -636.05789

Iteration 4: log pseudolikelihood = -635.63857

Iteration 5: log pseudolikelihood = -635.63667

Iteration 6: log pseudolikelihood = -635.63667

Logistic regression Number of obs = 2380

Wald chi2(10) = 265.96

Prob > chi2 = 0.0000

Log pseudolikelihood = -635.63667 Pseudo R2 = 0.2711

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

| Robust

deny | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

1.black | .6884231 .1821237 3.78 0.000 .3314673 1.045379

pi\_rat | 4.764416 1.329396 3.58 0.000 2.158848 7.369985

hse\_inc | -.1088114 1.294986 -0.08 0.933 -2.646938 2.429315

ltv\_med | .463525 .1600764 2.90 0.004 .149781 .777269

ltv\_high | 1.494764 .3242173 4.61 0.000 .8593095 2.130218

ccred | .2903017 .0388286 7.48 0.000 .2141991 .3664043

mcred | .2790178 .1376277 2.03 0.043 .0092724 .5487631

pubrec | 1.225797 .2030504 6.04 0.000 .8278253 1.623768

denpmi | 4.548166 .5744167 7.92 0.000 3.42233 5.674002

selfemp | .6661288 .2133542 3.12 0.002 .2479623 1.084295

\_cons | -5.707384 .4834338 -11.81 0.000 -6.654896 -4.759871

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

. eststo Logit\_2;

. quietly estadd margins black, atmeans;

. mat m = e(margins\_b);

. estadd scalar prob\_white = m[1,1];

added scalar:

e(prob\_white) = .0702292

. estadd scalar prob\_black = m[1,2];

added scalar:

e(prob\_black) = .13070374

. estadd scalar prob\_diff = m[1,2] - m[1,1];

added scalar:

e(prob\_diff) = .06047453

. margins black, atmeans vsquish;

Adjusted predictions Number of obs = 2380

Model VCE : Robust

Expression : Pr(deny), predict()

at : 0.black = .857563 (mean)

1.black = .142437 (mean)

pi\_rat = .3308136 (mean)

hse\_inc = .2553461 (mean)

ltv\_med = .3743697 (mean)

ltv\_high = .0323529 (mean)

ccred = 2.116387 (mean)

mcred = 1.721008 (mean)

pubrec = .0735294 (mean)

denpmi = .0201681 (mean)

selfemp = .1163866 (mean)

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

| Delta-method

| Margin Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

black |

0 | .0702292 .0061475 11.42 0.000 .0581803 .0822781

1 | .1307037 .0200064 6.53 0.000 .0914919 .1699156

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

We run a probit model in the same way using the –probit- command

. \*\* col(3) - Probit with robust standard errors;

. probit deny i.black `controls', r;

Iteration 0: log pseudolikelihood = -872.0853

Iteration 1: log pseudolikelihood = -643.92086

Iteration 2: log pseudolikelihood = -636.87532

Iteration 3: log pseudolikelihood = -636.84706

Iteration 4: log pseudolikelihood = -636.84706

Probit regression Number of obs = 2380

Wald chi2(10) = 293.99

Prob > chi2 = 0.0000

Log pseudolikelihood = -636.84706 Pseudo R2 = 0.2697

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

| Robust

deny | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

1.black | .3891261 .0977879 3.98 0.000 .1974653 .5807869

pi\_rat | 2.441787 .6092459 4.01 0.000 1.247687 3.635887

hse\_inc | -.1846764 .6753349 -0.27 0.785 -1.508308 1.138956

ltv\_med | .2139677 .0815119 2.62 0.009 .0542073 .3737281

ltv\_high | .7910831 .1797445 4.40 0.000 .4387903 1.143376

ccred | .1546224 .0209979 7.36 0.000 .1134672 .1957776

mcred | .1477082 .072594 2.03 0.042 .0054266 .2899899

pubrec | .6974496 .1153109 6.05 0.000 .4714445 .9234548

denpmi | 2.556796 .2983301 8.57 0.000 1.97208 3.141512

selfemp | .3586252 .1126041 3.18 0.001 .1379252 .5793252

\_cons | -3.040568 .2300051 -13.22 0.000 -3.49137 -2.589766

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

. quietly estadd margins black, atmeans;

. mat m = e(margins\_b);

. estadd scalar prob\_white = m[1,1];

added scalar:

e(prob\_white) = .07377839

. estadd scalar prob\_black = m[1,2];

added scalar:

e(prob\_black) = .14477953

. estadd scalar prob\_diff = m[1,2] - m[1,1];

added scalar:

e(prob\_diff) = .07100114

. eststo Probit\_3;

. margins black, atmeans;

Adjusted predictions Number of obs = 2380

Model VCE : Robust

Expression : Pr(deny), predict()

at : 0.black = .857563 (mean)

1.black = .142437 (mean)

pi\_rat = .3308136 (mean)

hse\_inc = .2553461 (mean)

ltv\_med = .3743697 (mean)

ltv\_high = .0323529 (mean)

ccred = 2.116387 (mean)

mcred = 1.721008 (mean)

pubrec = .0735294 (mean)

denpmi = .0201681 (mean)

selfemp = .1163866 (mean)

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

| Delta-method

| Margin Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

black |

0 | .0737784 .0063951 11.54 0.000 .0612442 .0863126

1 | .1447795 .0209432 6.91 0.000 .1037316 .1858274

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

Now we add some more control variables

. \*\* col(4) - Probit with more controls;

. probit deny i.black `controls' single hischl probunmp, r;

Iteration 0: log pseudolikelihood = -872.0853

Iteration 1: log pseudolikelihood = -636.63836

Iteration 2: log pseudolikelihood = -628.6448

Iteration 3: log pseudolikelihood = -628.61368

Iteration 4: log pseudolikelihood = -628.61368

Probit regression Number of obs = 2380

Wald chi2(13) = 311.96

Prob > chi2 = 0.0000

Log pseudolikelihood = -628.61368 Pseudo R2 = 0.2792

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

| Robust

deny | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

1.black | .3710683 .0987537 3.76 0.000 .1775146 .564622

pi\_rat | 2.464318 .5992224 4.11 0.000 1.289864 3.638772

hse\_inc | -.3024734 .6753734 -0.45 0.654 -1.626181 1.021234

ltv\_med | .2157285 .0818919 2.63 0.008 .0552234 .3762336

ltv\_high | .7945421 .1808344 4.39 0.000 .4401132 1.148971

ccred | .1575922 .0210859 7.47 0.000 .1162647 .1989198

mcred | .1104117 .0757864 1.46 0.145 -.038127 .2589504

pubrec | .7021586 .1160344 6.05 0.000 .4747354 .9295818

denpmi | 2.585178 .2937658 8.80 0.000 2.009408 3.160949

selfemp | .3459452 .1147379 3.02 0.003 .121063 .5708273

single | .2294801 .0798497 2.87 0.004 .0729776 .3859826

hischl | -.612624 .2312135 -2.65 0.008 -1.065794 -.159454

probunmp | .0300305 .0181281 1.66 0.098 -.0055 .065561

\_cons | -2.574574 .3352743 -7.68 0.000 -3.231699 -1.917448

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

. quietly estadd margins black, atmeans;

. mat m = e(margins\_b);

. estadd scalar prob\_white = m[1,1];

added scalar:

e(prob\_white) = .07193624

. estadd scalar prob\_black = m[1,2];

added scalar:

e(prob\_black) = .13775686

. estadd scalar prob\_diff = m[1,2] - m[1,1];

added scalar:

e(prob\_diff) = .06582062

. eststo Probit\_4;

. test single hischl probunmp;

( 1) [deny]single = 0

( 2) [deny]hischl = 0

( 3) [deny]probunmp = 0

chi2( 3) = 17.54

Prob > chi2 = 0.0005

And some more control variables

. \*\* col(5) - Probit with even more controls;

. probit deny i.black `controls' single hischl probunmp mcred3 mcred4 ccred3 ccred4 ccred5 ccred6 condo, r;

Iteration 0: log pseudolikelihood = -872.0853

Iteration 1: log pseudolikelihood = -773.70204

Iteration 2: log pseudolikelihood = -772.39945

Iteration 3: log pseudolikelihood = -772.39881

Iteration 4: log pseudolikelihood = -772.39881

Probit regression Number of obs = 2380

Wald chi2(11) = 192.45

Prob > chi2 = 0.0000

Log pseudolikelihood = -772.39881 Pseudo R2 = 0.1143

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

| Robust

deny | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

1.black | .5757906 .0888732 6.48 0.000 .4016024 .7499789

single | .2524522 .0761641 3.31 0.001 .1031734 .401731

hischl | -.5280365 .2290647 -2.31 0.021 -.976995 -.0790779

probunmp | .0439994 .0158903 2.77 0.006 .012855 .0751439

mcred3 | .2867383 .2287193 1.25 0.210 -.1615434 .7350199

mcred4 | .1341729 .3070509 0.44 0.662 -.4676359 .7359816

ccred3 | .4451225 .141355 3.15 0.002 .1680719 .7221732

ccred4 | .4928747 .1712032 2.88 0.004 .1573227 .8284268

ccred5 | .6297142 .1122034 5.61 0.000 .4097996 .8496289

ccred6 | .9035498 .1070117 8.44 0.000 .6938108 1.113289

condo | -.0572259 .0842932 -0.68 0.497 -.2224376 .1079857

\_cons | -1.270517 .2475202 -5.13 0.000 -1.755647 -.7853859

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

. eststo Probit\_5;

. test single hischl probunmp;

( 1) [deny]single = 0

( 2) [deny]hischl = 0

( 3) [deny]probunmp = 0

chi2( 3) = 24.32

Prob > chi2 = 0.0000

. test mcred3 mcred4 ccred3 ccred4 ccred5 ccred6;

( 1) [deny]mcred3 = 0

( 2) [deny]mcred4 = 0

( 3) [deny]ccred3 = 0

( 4) [deny]ccred4 = 0

( 5) [deny]ccred5 = 0

( 6) [deny]ccred6 = 0

chi2( 6) = 98.36

Prob > chi2 = 0.0000

. test condo;

( 1) [deny]condo = 0

chi2( 1) = 0.46

Prob > chi2 = 0.4972

And we run with an interaction term

. \*\* col(6) - Probit with interaction;

. probit deny i.black `controls'

> selfemp single hischl probunmp blk\_pi blk\_hse;

note: selfemp omitted because of collinearity

Iteration 0: log likelihood = -872.0853

Iteration 1: log likelihood = -636.10667

Iteration 2: log likelihood = -628.37393

Iteration 3: log likelihood = -628.33217

Iteration 4: log likelihood = -628.33216

Probit regression Number of obs = 2380

LR chi2(15) = 487.51

Prob > chi2 = 0.0000

Log likelihood = -628.33216 Pseudo R2 = 0.2795

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

deny | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

1.black | .2462823 .3938629 0.63 0.532 -.5256748 1.018239

pi\_rat | 2.571702 .5854259 4.39 0.000 1.424288 3.719116

hse\_inc | -.5380259 .7255947 -0.74 0.458 -1.960165 .8841136

ltv\_med | .2158857 .0826142 2.61 0.009 .0539649 .3778065

ltv\_high | .7881879 .1761974 4.47 0.000 .4428474 1.133529

ccred | .1578707 .021675 7.28 0.000 .1153884 .2003529

mcred | .1114918 .0754066 1.48 0.139 -.0363023 .259286

pubrec | .7047677 .1198014 5.88 0.000 .4699612 .9395742

denpmi | 2.590097 .2858882 9.06 0.000 2.029766 3.150427

selfemp | .3475272 .1130931 3.07 0.002 .1258687 .5691856

selfemp | 0 (omitted)

single | .2255941 .0801531 2.81 0.005 .068497 .3826912

hischl | -.6198276 .239398 -2.59 0.010 -1.089039 -.1506162

probunmp | .0297126 .018134 1.64 0.101 -.0058293 .0652546

blk\_pi | -.5792648 1.378903 -0.42 0.674 -3.281865 2.123335

blk\_hse | 1.231736 1.663421 0.74 0.459 -2.028509 4.49198

\_cons | -2.543412 .3276657 -7.76 0.000 -3.185625 -1.901199

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

. quietly estadd margins black, atmeans;

. mat m = e(margins\_b);

. quietly estadd scalar prob\_white = m[1,1];

. quietly estadd scalar prob\_black = m[1,2];

. quietly estadd scalar prob\_diff = m[1,2] - m[1,1];

. eststo Probit\_inter;

. test single hischl probunmp;

( 1) [deny]single = 0

( 2) [deny]hischl = 0

( 3) [deny]probunmp = 0

chi2( 3) = 16.63

Prob > chi2 = 0.0008

. test 1.black blk\_pi blk\_hse;

( 1) [deny]1.black = 0

( 2) [deny]blk\_pi = 0

( 3) [deny]blk\_hse = 0

chi2( 3) = 14.87

Prob > chi2 = 0.0019

. test blk\_pi blk\_hse;

( 1) [deny]blk\_pi = 0

( 2) [deny]blk\_hse = 0

chi2( 2) = 0.56

Prob > chi2 = 0.7561

Now we show a summary table using esttab

. esttab LPM Logit\_2 Probit\_3 Probit\_4 Probit\_5 Probit\_inter, stats(prob\_white prob\_black prob\_diff) mtitle;

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

(1) (2) (3) (4) (5) (6)

LPM Logit\_2 Probit\_3 Probit\_4 Probit\_5 Probit\_inter

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

main

0.black 0 0 0 0 0 0

(.) (.) (.) (.) (.) (.)

1.black 0.0837\*\*\* 0.688\*\*\* 0.389\*\*\* 0.371\*\*\* 0.363\*\*\* 0.246

(3.71) (3.78) (3.98) (3.76) (3.64) (0.63)

pi\_rat 0.449\*\*\* 4.764\*\*\* 2.442\*\*\* 2.464\*\*\* 2.622\*\*\* 2.572\*\*\*

(3.95) (3.58) (4.01) (4.11) (4.30) (4.39)

hse\_inc -0.0480 -0.109 -0.185 -0.302 -0.502 -0.538

(-0.44) (-0.08) (-0.27) (-0.45) (-0.72) (-0.74)

ltv\_med 0.0314\* 0.464\*\* 0.214\*\* 0.216\*\* 0.215\*\* 0.216\*\*

(2.47) (2.90) (2.62) (2.63) (2.58) (2.61)

ltv\_high 0.189\*\*\* 1.495\*\*\* 0.791\*\*\* 0.795\*\*\* 0.836\*\*\* 0.788\*\*\*

(3.77) (4.61) (4.40) (4.39) (4.59) (4.47)

ccred 0.0308\*\*\* 0.290\*\*\* 0.155\*\*\* 0.158\*\*\* 0.344\*\* 0.158\*\*\*

(6.71) (7.48) (7.36) (7.47) (3.25) (7.28)

mcred 0.0209 0.279\* 0.148\* 0.110 0.162 0.111

(1.85) (2.03) (2.03) (1.46) (1.59) (1.48)

pubrec 0.197\*\*\* 1.226\*\*\* 0.697\*\*\* 0.702\*\*\* 0.717\*\*\* 0.705\*\*\*

(5.65) (6.04) (6.05) (6.05) (6.13) (5.88)

denpmi 0.702\*\*\* 4.548\*\*\* 2.557\*\*\* 2.585\*\*\* 2.589\*\*\* 2.590\*\*\*

(15.56) (7.92) (8.57) (8.80) (8.68) (9.06)

selfemp 0.0598\*\* 0.666\*\* 0.359\*\* 0.346\*\* 0.342\*\* 0

(2.92) (3.12) (3.18) (3.02) (2.97) (.)

single 0.229\*\* 0.230\*\* 0.226\*\*

(2.87) (2.71) (2.81)

hischl -0.613\*\* -0.604\* -0.620\*\*

(-2.65) (-2.55) (-2.59)

probunmp 0.0300 0.0280 0.0297

(1.66) (1.57) (1.64)

mcred3 -0.107

(-0.37)

mcred4 -0.383

(-0.90)

ccred3 -0.226

(-0.94)

ccred4 -0.251

(-0.75)

ccred5 -0.789

(-1.94)

ccred6 -0.905

(-1.78)

condo -0.0550

(-0.59)

blk\_pi -0.579

(-0.42)

blk\_hse 1.232

(0.74)

\_cons -0.183\*\*\* -5.707\*\*\* -3.041\*\*\* -2.575\*\*\* -2.896\*\*\* -2.543\*\*\*

(-6.61) (-11.81) (-13.22) (-7.68) (-7.47) (-7.76)

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

prob\_white 0.0702 0.0738 0.0719 0.0704 0.0743

prob\_black 0.131 0.145 0.138 0.134 0.115

prob\_diff 0.0605 0.0710 0.0658 0.0631 0.0411

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

t statistics in parentheses

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001