Lending Mortage Analysis - Modeling

Exploratory Data Analysis

Marjorie Blanco, Joe Thompson, Haodi Tu

Data Discription

The data set includes the following variables:

- APPROVE = 1 if mortgage loan was approved, = 0 otherwise
- GDLIN = 1 if credit history meets guidelines, = 0 otherwise
- LOANPRC = loan amount/purchase price
- OBRAT = other obligations as a percent of total income
- MALE = 1 if male, = 0 otherwise
- MARRIED = 1 if married, = 0 otherwise
- BLACK = 1 if black, = 0 otherwise
- HISPAN = 1 if Hispanic, = 0 otherwise

Logistic Model

$$log(p/1-p) = \beta_0 + \beta_1 * GDLIN + \beta_2 * OBRAT + \beta_3 * BLACK + \beta_4 * HISPAN + \beta_5 * LOANPRC$$

```
##
## Call:  glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC,
##     family = "binomial", data = data)
##
## Coefficients:
## (Intercept)        GDLIN1         OBRAT         BLACK1       HISPAN1
##     1.53274       3.73675      -0.03122      -0.91714      -0.82664
##     LOANPRC
##    -0.01670
##
## Degrees of Freedom: 1887 Total (i.e. Null);  1882 Residual
## Null Deviance:        1391
## Residual Deviance: 902.5      AIC: 914.5
```

For every one unit change in `OBRAT`, the log odds of loan approval (versus non loan approval) decreases by 0.0312188.

For every one unit change in `LOANPRC`, the log odds of loan approval (versus non loan approval) decreases by 0.0167011.

The log odds of loan approval for applicants that meet credit guidelines increases by 3.7367457.

The log odds of loan approval for Black applicants decreases by 0.9171449.

The log odds of loan approval for Hispanic applicants decreases by 0.8266367.

For example, for a black person whose credit history meets guidline (GDLIN = 1), loan amount price is 100 (LOANPRC = 100) and other obligations as a percent of total income is none (OBRAT = 0), the log odds of loan approval is 93.6%

CIs using profiled log-likelihood

```
##                    2.5 %        97.5 %
## (Intercept)  0.19416207  2.937219042
## GDLIN1       3.31229466  4.182112724
## OBRAT       -0.05333480 -0.009301146
## BLACK1      -1.39135036 -0.426283068
## HISPAN1     -1.43764172 -0.165924717
## LOANPRC     -0.03156216 -0.002916465
```

CIs using standard errors

```
##                    2.5 %        97.5 %
## (Intercept)  0.16272694  2.902760759
## GDLIN1       3.30270694  4.170784503
## OBRAT       -0.05325850 -0.009179045
## BLACK1      -1.39893109 -0.435358718
## HISPAN1     -1.46068910 -0.192584237
## LOANPRC     -0.03101509 -0.002387147
```
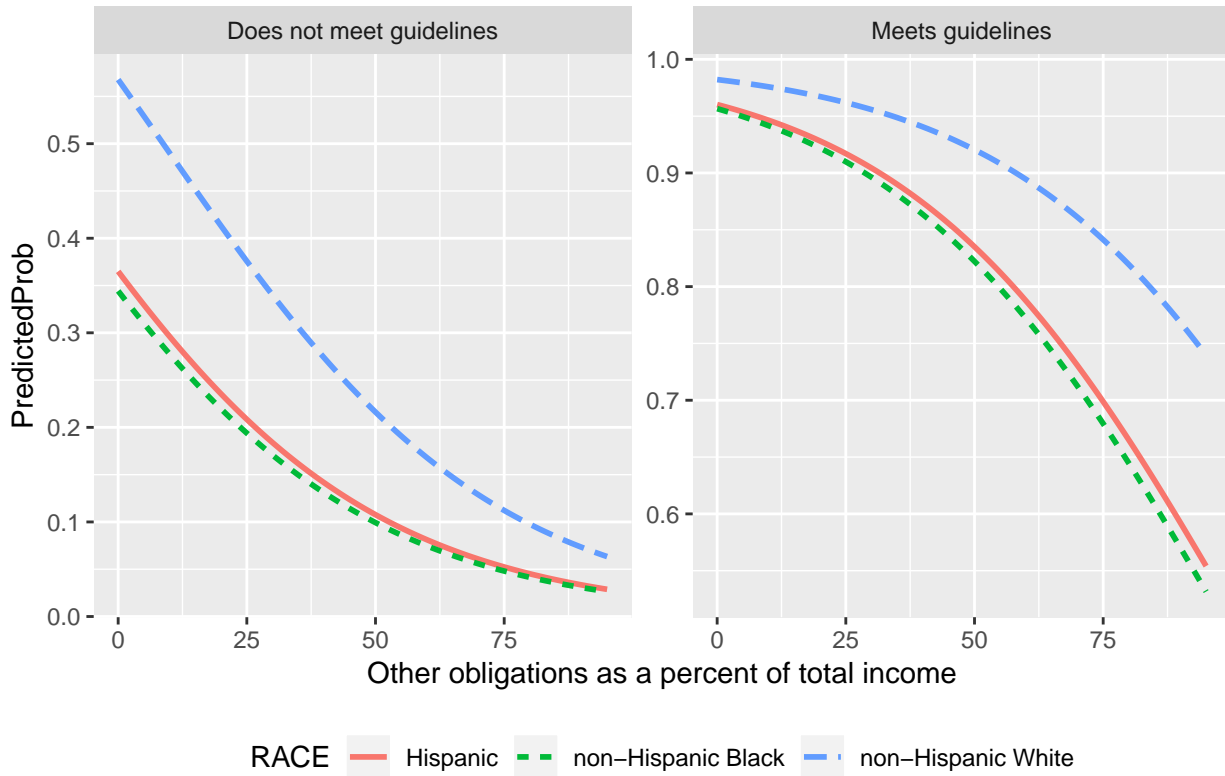
Odds ratios only

```
## (Intercept)      GDLIN1       OBRAT       BLACK1      HISPAN1      LOANPRC
##   4.6308658   41.9612143   0.9692635   0.3996585   0.4375183   0.9834376
```

Odds ratios and 95% CI
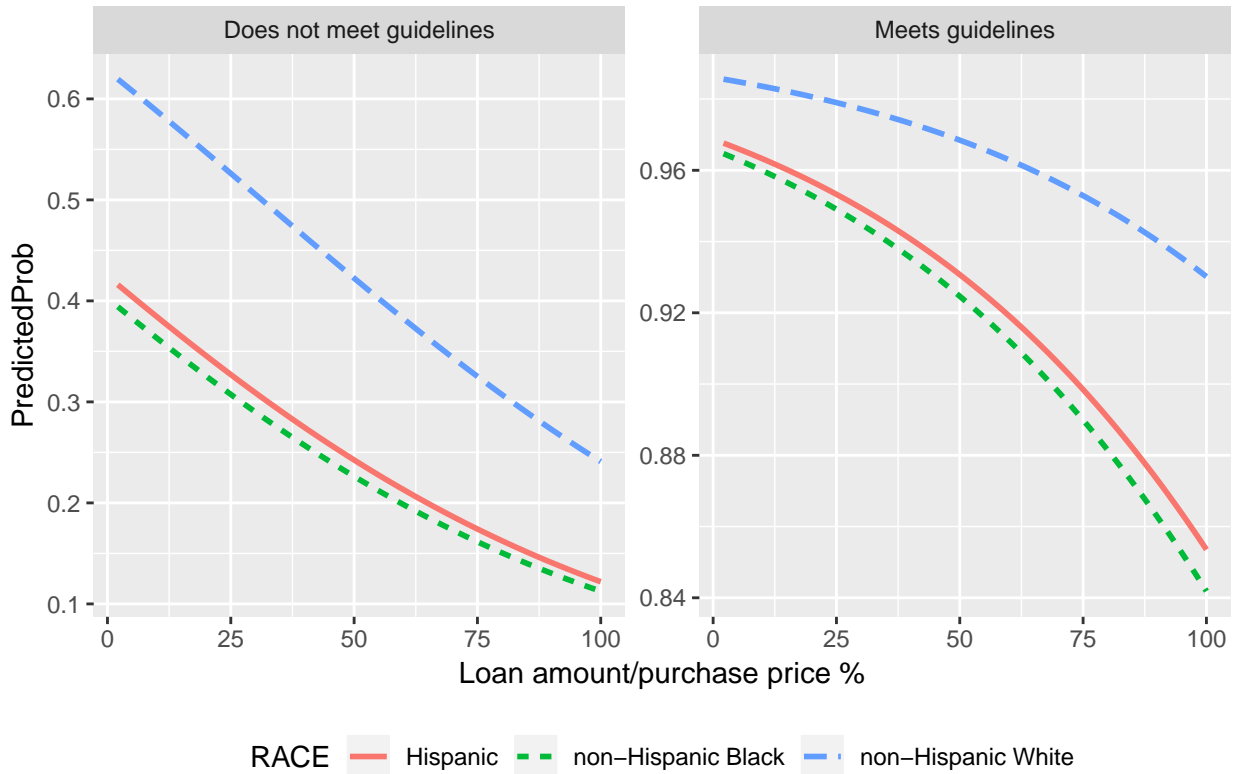
```
##                    OR        2.5 %       97.5 %
## (Intercept)  4.6308658   1.2142931  18.8633152
## GDLIN1      41.9612143  27.4480373  65.5040992
## OBRAT        0.9692635   0.9480625   0.9907420
## BLACK1       0.3996585   0.2487392   0.6529315
## HISPAN1      0.4375183   0.2374872   0.8471100
## LOANPRC      0.9834376   0.9689307   0.9970878
```

GDLIN	OBRAT	BLACK	HISPAN	LOANPRC	fit	PredictedProb
0	32.35767	1	0	75.44245	-1.6545412	0.1604961
1	32.35767	1	0	75.44245	2.0822046	0.8891615
0	32.35767	0	1	75.44245	-1.5640329	0.1730687
1	32.35767	0	1	75.44245	2.1727128	0.8977722
0	32.35767	0	0	75.44245	-0.7373963	0.3235738
1	32.35767	0	0	75.44245	2.9993495	0.9525447

Predicted probabilities (LOANPRC = 75.44245%)

Does not meet guidelines / **Meets guidelines**

PredictedProb

Other obligations as a percent of total income

RACE — Hispanic --- non−Hispanic Black --- non−Hispanic White

Predicted probabilities (OBRAT = 32.35767)

Does not meet guidelines / **Meets guidelines**

PredictedProb

Loan amount/purchase price %

RACE — Hispanic --- non−Hispanic Black --- non−Hispanic White

Probit Model

$$Probit(approve) = \beta_0 + \beta_1 * GDLIN + \beta_2 * OBRAT + \beta_3 * BLACK + \beta_4 * HISPAN + \beta_5 * LOANPRC$$

```
##
## Call:
## glm(formula = APPROVE ~ GDLIN + OBRAT + BLACK + HISPAN + LOANPRC,
##     family = binomial(link = "probit"), data = data)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.8688   0.2496   0.3144   0.3604   2.1669
##
## Coefficients:
##              Estimate Std. Error z value           Pr(>|z|)
## (Intercept)  0.583275   0.341346   1.709           0.087497 .
## GDLIN1       2.161541   0.124153  17.410 < 0.0000000000000002 ***
## OBRAT       -0.014974   0.005721  -2.617           0.008858 **
## BLACK1      -0.473416   0.129376  -3.659           0.000253 ***
## HISPAN1     -0.422091   0.168680  -2.502           0.012338 *
## LOANPRC     -0.007808   0.003404  -2.294           0.021781 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1391.24  on 1887  degrees of freedom
## Residual deviance:  902.36  on 1882  degrees of freedom
## AIC: 914.36
##
## Number of Fisher Scoring iterations: 6
```

For every one unit change in `OBRAT`, the probit odds of loan approval (versus non loan approval) decreases by 0.0149739.

For every one unit change in `LOANPRC`, the probit odds of loan approval (versus non loan approval) decreases by 0.0078082.

The probit odds of loan approval for applicants that meet credit guidelines increases by 2.1615407.
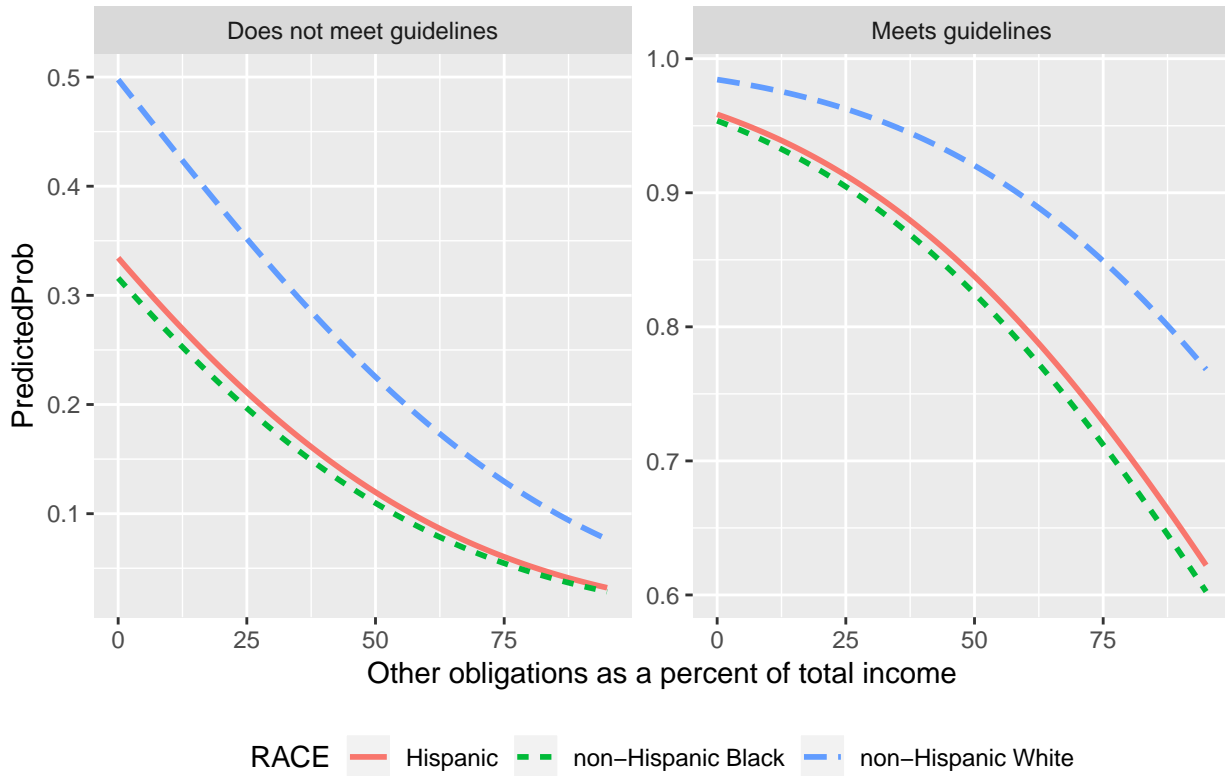
The probit odds of loan approval for Black applicants decreases by 0.4734163.

The probit odds of loan approval for Hispanic applicants decreases by 0.4220909.

For example, for a black person whose credit history meets guidline (GDLIN = 1), loan amount price is 100 (LOANPRC = 100) and other obligations as a percent of total income is none (OBRAT = 0), the probit odds of loan approval is 93.2%

GDLIN	OBRAT	BLACK	HISPAN	LOANPRC	fit
0	32.35767	1	0	75.44245	0.3158938
1	32.35767	1	0	75.44245	0.3105874
0	32.35767	0	1	75.44245	0.3053201
1	32.35767	0	1	75.44245	0.3000927
0	32.35767	0	0	75.44245	0.2949063
1	32.35767	0	0	75.44245	0.2897615

Predicted probabilities (LOANPRC = 75.44245%)

Predicted probabilities (OBRAT = 32.35767)