***Mortgage Lending in Boston: Interpreting HMDA Data***

*Working with equations 11.1, 11.3, 11.7, 11.8, 11.10*:

To make the data on these simple regressions a little more readable we could output it using the esttab command.

Mortgage Denial Regressed On P/I Ratio and Race

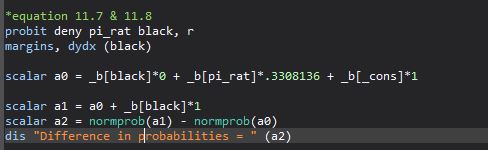
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |
|  | LPM | LPM | Probit | Probit | Logit |
|  |  |  |  |  |  |
| black |  | 0.177\*\*\* |  | 0.708\*\*\* | 1.273\*\*\* |
|  |  | (0.0249) |  | (0.0832) | (0.146) |
|  |  |  |  |  |  |
| Pi  \_rat | 0.604\*\*\* | 0.559\*\*\* | 2.968\*\*\* | 2.742\*\*\* | 5.370\*\*\* |
|  | (0.0985) | (0.0887) | (0.465) | (0.444) | (0.963) |
|  |  |  |  |  |  |
| Constant | -0.0799\* | -0.0905\*\* | -2.194\*\*\* | -2.259\*\*\* | -4.126\*\*\* |
|  | (0.0320) | (0.0286) | (0.165) | (0.159) | (0.346) |
| Observations | 2380 | 2380 | 2380 | 2380 | 2380 |

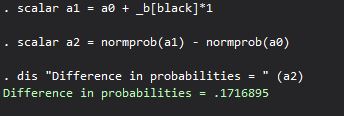
Standard errors in parentheses

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

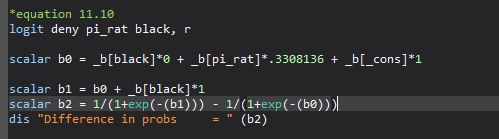
Equations 11.1 and 11.3 are straightforward linear regressions. We see from equation 11.3, which is regression (2) that a black individual’s chance of mortgage denial is 17.7 percentage points higher than a white individual, which is statistically significant at an alpha of .001. Likewise, with the logit and probit models, since they are non-linear we can’t interpret the coefficients as we do in a linear model. For example, in the probit models, the coefficients do not show the marginal effects in the probability of denial, but rather the change in z-value associated with changes in X. As a result, with non-linear models, such as logits and probits, the best way to interpret the coefficients is to compute the predicted probability change for one or more values of the regressors.

We can replicate similar results with the two simple probit models in 11.7 and 11.8. I explain how this is done in more detail below in the subsequent examples used in replicating table 11.2. However, in short, we essentially take specific values for all of the regressors except for race, then we compute the difference in predicted probabilities of denial for a black applicant and white applicant. For those specific values we will simply use the sample average values of those regressors.





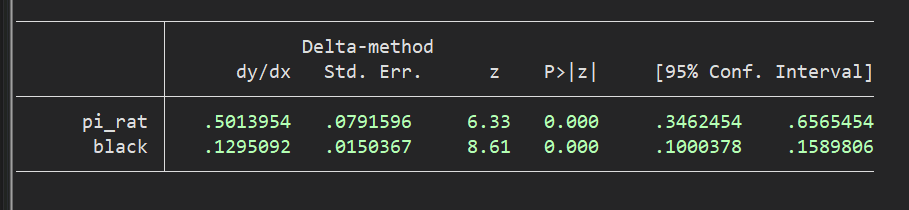
Based on this simple probit model regressing mortgage denial rates on P/I Ratio and black, we see similar results. The difference in expected probabilities of denial between a white and black applicant is 17.17 percentage points, which is almost identical to the LPM race differential above.

We can extend this to the simple logit model in 11.10.



Based on this simple logit model (11.10), the difference in expected probabilities of denial between a white and black applicant is 16.71 percentage points. This differential is almost identical to the models above.

Likewise, we could find the marginal effects of each regressors, or how much the conditional probability of the dependent variable changes when changing the value of a regressor. For example, in the simple Probit regression in equation 11.8 we could just run the regression followed by the margins command.



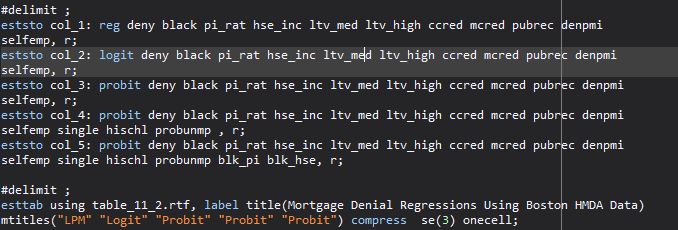
A one-unit increase in P/I ratio would increase the predicted probability of denial by 50 percentage points. Likewise, being black increases your chance of denial by 13 percentage points. Both are statistically significant at an alpha of .05.

Replicating the first half of table 11.2 is relatively straightforward. I replicated the 5 regression functions and used the eststo command to store the estimation results for later tabulation in the esttab table. The 5 regression functions, respectively, are:

1. LPM – =
2. Logit – =
3. Probit – =
4. Probit – =
5. Probit – =

The first three regression functions are the same, however, the model used in each is different. Regression (4) extends regression (3) by adding the regressors single, high school diploma and unemployment rate. Regression (5) extends regression (4) by adding the interaction effects for black \* P/I ratio and black \* housing expense-to-income ratio.

Likewise, we can store each of the regressions using the eststo command and then display all the data in an esttab table in similar formatting to the textbook.



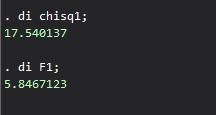
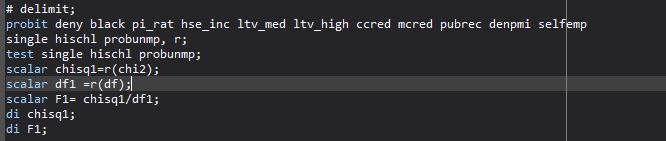
Mortgage Denial Regressions Using Boston HMDA Data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |
|  | LPM | Logit | Probit | Probit | Probit |
|  |  |  |  |  |  |
| black | 0.0837\*\*\* (0.023) | 0.688\*\*\* (0.182) | 0.389\*\*\* (0.098) | 0.371\*\*\* (0.099) | 0.246 (0.448) |
|  |  |  |  |  |  |
| P/I ratio | 0.449\*\*\* (0.114) | 4.764\*\*\* (1.329) | 2.442\*\*\* (0.609) | 2.464\*\*\* (0.599) | 2.572\*\*\* (0.662) |
|  |  |  |  |  |  |
| monthly housing expenses/income | -0.0480 (0.110) | -0.109 (1.295) | -0.185 (0.675) | -0.302 (0.675) | -0.538 (0.743) |
|  |  |  |  |  |  |
| medium loan-to-value ratio | 0.0314\* (0.013) | 0.464\*\* (0.160) | 0.214\*\* (0.082) | 0.216\*\* (0.082) | 0.216\*\* (0.082) |
|  |  |  |  |  |  |
| high loan-to-value ratio | 0.189\*\*\* (0.050) | 1.495\*\*\* (0.324) | 0.791\*\*\* (0.180) | 0.795\*\*\* (0.181) | 0.788\*\*\* (0.181) |
|  |  |  |  |  |  |
| consumer credit score | 0.0308\*\*\* (0.005) | 0.290\*\*\* (0.039) | 0.155\*\*\* (0.021) | 0.158\*\*\* (0.021) | 0.158\*\*\* (0.021) |
|  |  |  |  |  |  |
| mortgage credit score | 0.0209 (0.011) | 0.279\* (0.138) | 0.148\* (0.073) | 0.110 (0.076) | 0.111 (0.076) |
|  |  |  |  |  |  |
| public bad credit record | 0.197\*\*\* (0.035) | 1.226\*\*\* (0.203) | 0.697\*\*\* (0.115) | 0.702\*\*\* (0.116) | 0.705\*\*\* (0.116) |
|  |  |  |  |  |  |
| denied mortgage insurance | 0.702\*\*\* (0.045) | 4.548\*\*\* (0.574) | 2.557\*\*\* (0.298) | 2.585\*\*\* (0.294) | 2.590\*\*\* (0.294) |
|  |  |  |  |  |  |
| self employed | 0.0598\*\* (0.021) | 0.666\*\* (0.213) | 0.359\*\* (0.113) | 0.346\*\* (0.115) | 0.348\*\* (0.115) |
|  |  |  |  |  |  |
| single |  |  |  | 0.229\*\* (0.080) | 0.226\*\* (0.080) |
|  |  |  |  |  |  |
| high school diploma |  |  |  | -0.613\*\* (0.231) | -0.620\*\* (0.231) |
|  |  |  |  |  |  |
| unemployment rate |  |  |  | 0.0300 (0.018) | 0.0297 (0.018) |
|  |  |  |  |  |  |
| black \* P/I ratio |  |  |  |  | -0.579 (1.473) |
|  |  |  |  |  |  |
| black \* housing expense-to-income ratio |  |  |  |  | 1.232 (1.694) |
|  |  |  |  |  |  |
| Constant | -0.183\*\*\* (0.028) | -5.707\*\*\* (0.483) | -3.041\*\*\* (0.230) | -2.575\*\*\* (0.335) | -2.543\*\*\* (0.349) |
| Observations | 2380 | 2380 | 2380 | 2380 | 2380 |

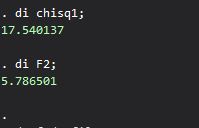
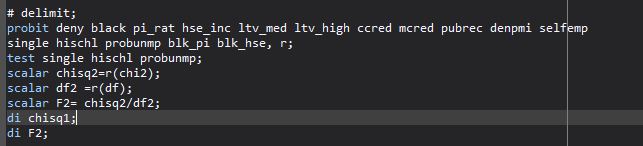
Standard errors in parentheses

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

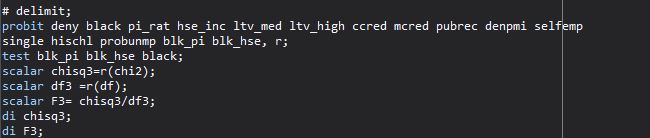
Now onto replicating the second portion of table 11.2, involving the f-statistics. Replicating the test for applicant single, high school diploma and unemployment rate in regression (4) in the textbook:

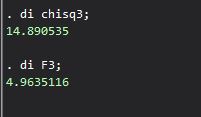


Replicating that same test for applicant single, high school diploma and unemployment rate but for regression (6) in the textbook. The only difference here is that this regression model includes the two interaction effects, so we need to include those in the regression function and re-run the test.

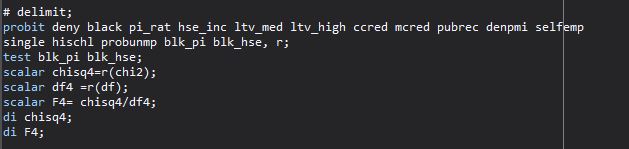


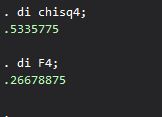
The following is the test for race interactions and black in regression (6) of the textbook. We take the same regression as above but just test different variables. In this case those variables to test are blk\_pi, blk\_hse and black.





The following is testing race interactions only, in regression (6) of the textbook. We simply just remove the black variable from the test.





Replicating the difference in predicted probabilities:

*For the LPM regression (1) in the textbook:*

To replicate the difference in probabilities of denial between a white and black applicant for the LPM, we simply look at the coefficient on black in regression (1), which is .084. In other words, a black applicant is 8.4 percentage points more likely to have his/her mortgage application denied than a white person. Likewise, this is statistically significant at an alpha of .01 or 99%.

However, for the logits and probits, since these models are non-linear, specific values for all of the regressors need to be chosen to compute the difference in predicted probabilities of denial for a black applicant and white applicant. One way to do this is to consider an “average” applicant who has the sample averages of all the regressors except for race. We then compute the predicted probability difference of mortgage denial for this average applicant based on whether he/she is white or black, which is what is done in the last row of 11.2 in the textbook.

In terms of Stata, we first need to compute the means of all the co-variables. Then we input the means of the regressors for that specific regression and multiply them by their respective coefficients. Adding up all the terms, we then use either the logistic distribution for logits or the normal distribution for probits, to find the respective probabilities.

For logit regressions with multiple regressors:

Pr() = F () =

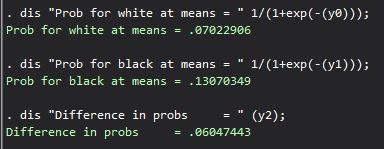
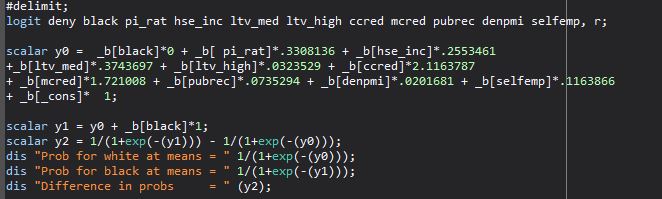
For probit regressions with multiple regressors:

Pr() = Φ ()

We simply take the difference for a white and black applicant to determine the predicted probability difference. In the above cases the X values are simply the means for those regressors. We can find the mean for all the co-variables by just running the sum command.

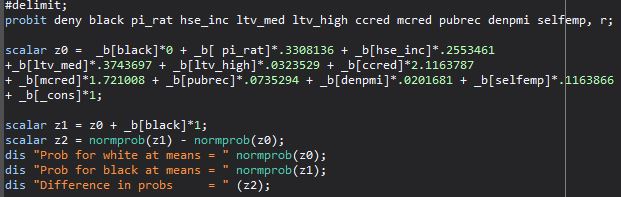


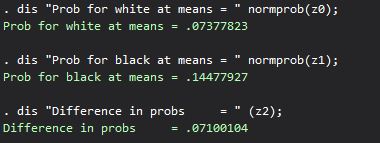
*For the logit regression (2) in the textbook:*



The estimated difference in denial probability for a white vs black applicant in the logit regression (2) is .0605 = 6 percentage points. Which is what is shown in table 11.2.

*For the probit regression (3) in the textbook:*

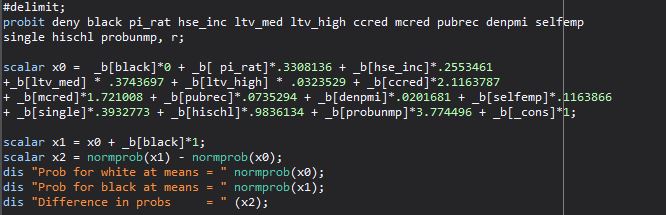


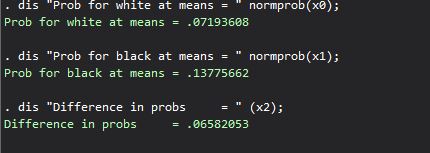


Our probability difference in denial of a mortgage between a black vs white applicant in the probit regression (3) is .071001 = 7.1 percentage points.

*For the probit regression (4) in the textbook:*

Here the regression function extends to include the variables – single, high school diploma and unemployement rate. We need to include these regressors in the regression function and then multiply their coefficients by our “average” applicant, as we have been doing.

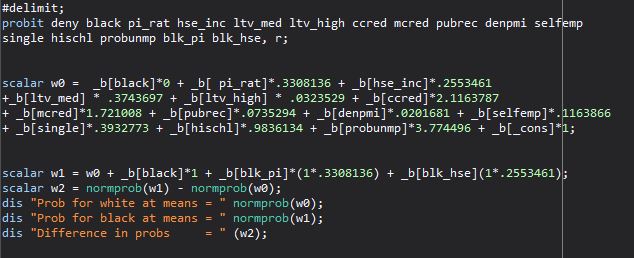


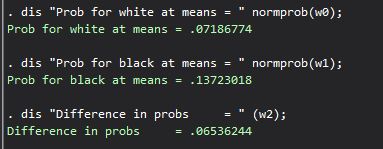


Our predicted probability difference between white and black applicants in the probit regression (4) is .0658 = 6.6 percentage points, as mentioned in the table 11.2.

*For the probit regression (6) in the textbook:*

This regression function extends the regression in (4) by adding the race interactions, black \* P/I Ratio and black \* housing expense-to-income ratio. Again, we need to include these interaction terms in the regression function and multiply the coefficients by the “average” for that variable.





Our predicted probability difference between white and black applicants in the probit regression (6) is .065= 6.5 percentage points, as mentioned in the table 11.2.

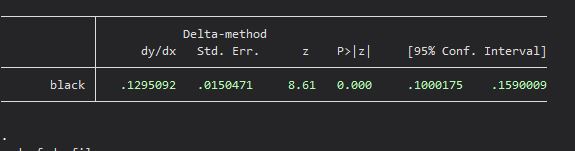
*Questions*

1. When dealing with the output of a probit regression, we can’t directly interpret the coefficients as we do in a linear model. Instead we need to interpret the marginal effects of the regressors, or how much the conditional probability of the dependent variable changes when we change the value of a regressor, holding all the other regressors constant. So in order to compare the results for the regressor black from regression (6) to that of equation 11.8, we need to find the marginal effects in each case.

We can simply use the margins command-

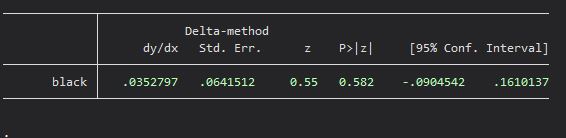
For the simple probit regression in equation 11.8:





In other words, when an applicant is black, his or her chance of mortgage denial increases by 12.95 percentage points, which is statistically significant.

For the probit regression (6):



In other words, in this regression model, when an applicant is black, his or her chance of mortgage denial increases by 3.5 percentage points, which is not statistically significant. Likewise, the estimated effect of being black on mortgage denial is a lot lower than in equation 11.8, which could suggest the presence of omitted variable bias in equation 11.8.

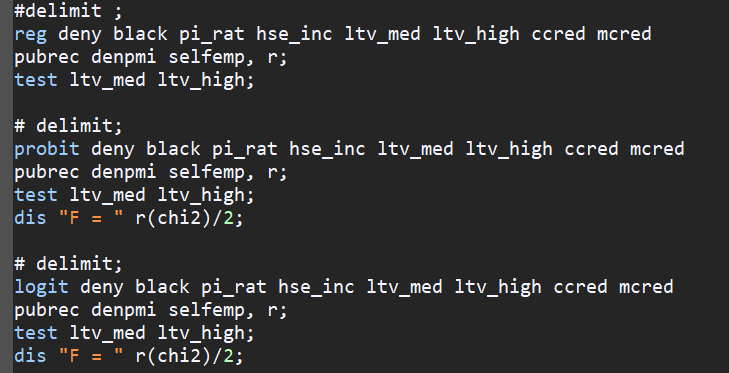
1. The regressions in (1) through (3) only differ in how the probability is modeled; an LPM, Logit and Probit. For the LPM in regression (1), we could simply use the coefficients to determine the marginal effects in the predicted probabilty in mortgage denial. As an example, being black increases your chances of denial by 8.4 percentage points, which is statistically significant at an alpha of .01. A one unit increase in P/I ratio increases the chance of denial by 45 percentage points, which is also statistically significant at an alpha of .01. Likewise, out of the nine variables, all but two are statistically significant at the 95% level, which suggests that loan officers have many significant factors they could consider when reviewing an applicant for a mortgage.

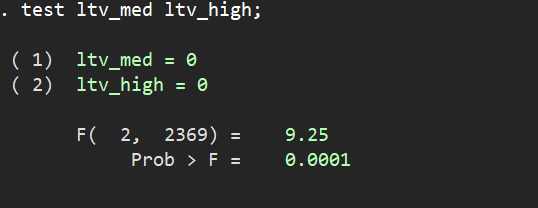
However, in the Logit and Probit models, since these are non-linear, we can’t interpret the coefficients as we do in th LPM. A better way to gauge the coefficients in these models is to compute the difference in predicted probabibilites between a white and black applicant at specific values for all the regressors except for race. As mentioned above, we could consider an ”average” applicant whose regressor values are simply the averages. When comparing these predicted racial differentials, all three models produce similar results. Regression (1) (LPM) estimates an 8.4 percentage point difference, Regression (2) (Logit) estimates a a 6.0 percentage point difference and Regression (3) (Probit) predicts a 7.1 percentage point difference.

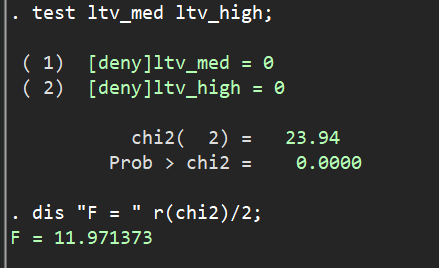
Another way to compare the 3 models is to simply look at the marginal effects each regressor has on the predicted probability of denial. For the LPM, this simply entails looking at the regression output and examining the coefficients. For the Logit and Probit model we need to take the extra step of computing those marginal effects by using the margins command in stata. This might give us a more specific look into how each regressor would affect the predicted probability, where as above, we were just looking at general racial differentials amongst the different models.

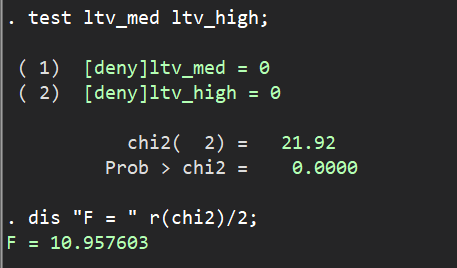
Does the type of model make a difference? Obviously the marginal effects between the three models are slightly different. The marginal effects in the logit and probit models are almost identical and these effects are only slightly different from the LPM. In terms of which model to use to make conclusions about our data, I think all three models give the same general conclusion about the data. Likewise, the statistical significance of all the regressors, except one, is constant throughtout the models. That one exception is the mortgage credit score, which is significant in the logit and probit but not significant in the LPM. Likewise, you could still come to the same conclusion in all the models, which is that there is racial bias in mortgage applications even when factoring other common variables such as credit score, ltv ratios, etc.

1. The way I would test the construct of loan-to-value ratio is by testing those variables in regressions (1), (2) and (3), as these regression functions have the same regressors and the only difference in these regressions are the way they are modeled.



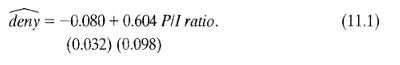


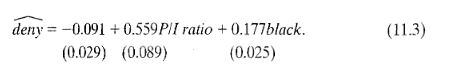


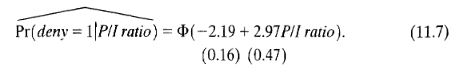


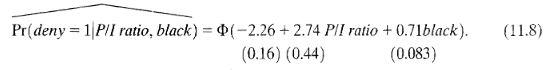
All the tests produce similar results; loan-to-value ratios (both med and high) are statistically significant. In other words, regression models that use this construct fit the data better than models without, which suggests that ltv ratios are a significant aspect that loan officers can take into account when reviewing a mortgage application.

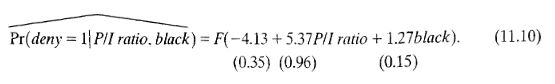
Equations that needed to replicated in the first part of the project.











Reference Questions

1. Compare the results for black in column 6 to those obtained in the simple probit model of Deny regressed on the P/I ratio and Black (*equation 11.8 in the text but make sure you estimate the effect for black in percentage points).* Make sure to discuss the units your estimates are in and its statistical significance.
2. Compare the results in column (1) to those in columns (2) and (3). Does the type of model make a difference? Explain.
3. How would you test the construct, “loan to value ratio”? Note they are dummy variables in these regressions?

Table 11.2 that needed to replicated.

