# Econ 717 Problem Set 1

Sarah Bass

February 16, 2022

## Question 1

I have dropped 65 observations from the data due to missing values for household income, marital status, education, and age.

## Question 2

Table 1. Linear Probability Model

| | (1) |
|---|---|
| VARIABLES | taken_new |
| client_age | -2.83e-05 |
| | (0.00216) |
| client_married | 0.0117 |
| | (0.0529) |
| client_education | -0.00369 |
| | (0.00412) |
| hh_size | -0.0113 |
| | (0.00931) |
| hh_income | 3.14e-06 |
| | (3.68e-06) |
| muslim | -0.00756 |
| | (0.0367) |
| hindu_sc_kat | -0.0275 |
| | (0.0526) |
| treated | 0.0426 |
| | (0.0347) |
| Constant | 0.199* |
| | (0.114) |
| Observations | 532 |
| R-squared | 0.008 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

First note that the constant term has the only statistically significant coefficient; all others are not statistically significantly different from 0. All of the coefficients are small in magnitude, which is consistent with our expectations for coefficients in the linear probability model since our outcome variable is binary. We can also observe that our constant term is roughly 0.199, which is roughly equivalent to the unconditional probability of taking out a new loan (0.17).

# Question 3

Table 2. Linear Probability Model, Robust Standard Errors

| | (1) | (2) |
|---|---|---|
| VARIABLES | taken_new | taken_new |
| client_age | -2.83e-05 | -2.83e-05 |
| | (0.00216) | (0.00227) |
| client_married | 0.0117 | 0.0117 |
| | (0.0529) | (0.0519) |
| client_education | -0.00369 | -0.00369 |
| | (0.00412) | (0.00410) |
| hh_size | -0.0113 | -0.0113 |
| | (0.00931) | (0.00928) |
| hh_income | 3.14e-06 | 3.14e-06 |
| | (3.68e-06) | (3.71e-06) |
| muslim | -0.00756 | -0.00756 |
| | (0.0367) | (0.0365) |
| hindu_sc_kat | -0.0275 | -0.0275 |
| | (0.0526) | (0.0510) |
| treated | 0.0426 | 0.0426 |
| | (0.0347) | (0.0335) |
| Constant | 0.199* | 0.199* |
| | (0.114) | (0.117) |
| Observations | 532 | 532 |
| R-squared | 0.008 | 0.008 |
| Robust S.E. | No | Yes |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Compared to the standard errors in Table 1, the robust standard errors in Table 2 are larger for client age, household income, and the constant term. For all other variables, the robust standard errors are smaller. Note, even with the smaller standard errors, the constant term remains the only statistically significant coefficient.

## Question 4

There are no observations with predicted probabilities of loan take-up that lie outside of the [0,1] interval.

## Question 5

Due to the small sample size, there are not sufficient observations to estimate the model by weighted least squares.

## Question 6

Table 3. Probit and Logit Models

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | taken_new | taken_new | taken_new |
| client_age | -2.83e-05 | 0.000154 | -0.000382 |
| | (0.00216) | (0.00856) | (0.0157) |
| client_married | 0.0117 | 0.0495 | 0.0931 |
| | (0.0529) | (0.214) | (0.388) |
| client_education | -0.00369 | -0.0146 | -0.0276 |
| | (0.00412) | (0.0166) | (0.0300) |
| hh_size | -0.0113 | -0.0476 | -0.0854 |
| | (0.00931) | (0.0379) | (0.0694) |
| hh_income | 3.14e-06 | 1.33e-05 | 2.23e-05 |
| | (3.68e-06) | (1.44e-05) | (2.53e-05) |
| muslim | -0.00756 | -0.0326 | -0.0533 |
| | (0.0367) | (0.147) | (0.263) |
| hindu_sc_kat | -0.0275 | -0.110 | -0.208 |
| | (0.0526) | (0.215) | (0.395) |
| treated | 0.0426 | 0.175 | 0.319 |
| | (0.0347) | (0.142) | (0.259) |
| Constant | 0.199* | -0.853* | -1.374 |
| | (0.114) | (0.459) | (0.835) |
| Observations | 532 | 532 | 532 |
| R-squared | 0.008 | | |
| Model | LPM | Probit | Logit |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The probit, logit, and LPM coefficients are not identical, and the probit and logit estimates are much closer to each other than they are to the LPM coefficients. This is because both probit and logit models use a CDF link function with normal and log-normal distributions, respectively. However, the LPM does not use a CDF for a link function, so the estimates differ some.

# Question 7

## Part A

Table 4. Mean Partial Derivative Estimates using dprobit

| | (1) |
|---|---|
| VARIABLES | taken_new |
| client_age | 3.82e-05 |
| | (0.00213) |
| client_married | 0.0121 |
| | (0.0511) |
| client_education | -0.00363 |
| | (0.00412) |
| hh_size | -0.0118 |
| | (0.00939) |
| hh_income | 3.30e-06 |
| | (3.57e-06) |
| muslim | -0.00804 |
| | (0.0360) |
| hindu_sc_kat | -0.0262 |
| | (0.0490) |
| treated | 0.0422 |
| | (0.0332) |
| Observations | 532 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Using the dprobit command in Stata, I find the the mean partial derivative is 0.0000382.

## Part B

When calculated analytically by hand, I find that the mean partial derivative is 0.0000382.

## Part C

When calculated numerically, I find that the mean partial derivative is 0.0000382.

## Part D

When calculated using the margins command, I find that the mean partial derivative is 0.0000382.

# Question 8

Table 5. Linear Probability Model with Age Polynomial

| | (1) | (2) |
|---|---|---|
| VARIABLES | taken_new | taken_new |
| client_age | -2.83e-05 | -0.493*** |
| | (0.00216) | (0.179) |
| client_married | 0.0117 | 0.0152 |
| | (0.0529) | (0.0553) |
| client_education | -0.00369 | -0.00315 |
| | (0.00412) | (0.00411) |
| hh_size | -0.0113 | -0.00889 |
| | (0.00931) | (0.00931) |
| hh_income | 3.14e-06 | 3.65e-06 |
| | (3.68e-06) | (3.67e-06) |
| muslim | -0.00756 | -0.0128 |
| | (0.0367) | (0.0365) |
| hindu_sc_kat | -0.0275 | -0.0333 |
| | (0.0526) | (0.0523) |
| treated | 0.0426 | 0.0465 |
| | (0.0347) | (0.0344) |
| client_age_2 | | 0.0199*** |
| | | (0.00688) |
| client_age_3 | | -0.000336*** |
| | | (0.000111) |
| client_age_4 | | 2.01e-06*** |
| | | (6.40e-07) |
| Constant | 0.199* | 4.518*** |
| | (0.114) | (1.656) |
| Observations | 532 | 532 |
| R-squared | 0.008 | 0.031 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

After adding higher order terms to this linear probability model, now all of the age variables are statistically significant. When I calculate the average partial numerically with respect to the age, I find that the partial is $3.35 \times 10e^{-5}$.Note

# Question 9

I have calculated an LRI value of 0.00899. A low value for the LRI indicates that $\hat{L} \approx L_0$ and little value is added from including additional covariates.

## Question 10

When I calculate the correct prediction rate for the probit model using a cutoff of 0.5, there are no observations with a predicted probability greater than 0.5, so everyone is predicted to not take out a loan. As a result, the 443 observations that did not take out a loan are correctly identified, but the 89 people who did take out a loan are incorrectly identified. In this case the correct prediction rate is 83.27.

When I change the cutoff to the fraction of the sample that takes out a loan (0.1673), I find a correct prediction rate of 0.5338. This cutoff value results in 216 "false positives" where the predicted probability is greater than 0.1673 but the individual did not take out a loan, as well as 32 people who had predicted probabilities below the cutoff, but they did take out a loan.

## Question 11

Table 6. Out-of-sample Probit Model

| | (1) | (2) |
|---|---|---|
| VARIABLES | taken_new | taken_new |
| client_age | 0.000154 | 0.00955 |
| | (0.00856) | (0.0133) |
| client_married | 0.0495 | 0.122 |
| | (0.214) | (0.317) |
| client_education | -0.0146 | 0.00200 |
| | (0.0166) | (0.0247) |
| hh_size | -0.0476 | 0.0451 |
| | (0.0379) | (0.0526) |
| hh_income | 1.33e-05 | 4.91e-06 |
| | (1.44e-05) | (2.00e-05) |
| muslim | -0.0326 | 0.00643 |
| | (0.147) | (0.212) |
| hindu_sc_kat | -0.110 | -0.506 |
| | (0.215) | (0.387) |
| treated | 0.175 | 0.0435 |
| | (0.142) | (0.203) |
| Constant | -0.853* | -1.708** |
| | (0.459) | (0.728) |
| Observations | 532 | 266 |
| Full Sample | Yes | Out of Sample |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Using a cutoff of 0.5, I estimated an out-of-sample correct prediction rate of 0.8383, however in this scenario no observations had predicted probabilities above the threshold, so everyone was predicted to not take out a loan. When I use a cutoff of 0.1673, I estimated an out-of-sample correct prediction rate of 0.5601.

## Question 12

Table 7. Probit Model with Married * Muslim Interaction

| | (1) | (2) |
|---|---|---|
| VARIABLES | taken_new | taken_new |
| client_age | 0.000154 | 0.000931 |
| | (0.00856) | (0.00868) |
| client_married | 0.0495 | 0.156 |
| | (0.214) | (0.281) |
| client_education | -0.0146 | -0.0154 |
| | (0.0166) | (0.0167) |
| hh_size | -0.0476 | -0.0502 |
| | (0.0379) | (0.0381) |
| hh_income | 1.33e-05 | 1.26e-05 |
| | (1.44e-05) | (1.45e-05) |
| muslim | -0.0326 | 0.206 |
| | (0.147) | (0.421) |
| hindu_sc_kat | -0.110 | -0.114 |
| | (0.215) | (0.215) |
| treated | 0.175 | 0.182 |
| | (0.142) | (0.142) |
| married_muslim | | -0.271 |
| | | (0.448) |
| Constant | -0.853* | -0.959* |
| | (0.459) | (0.494) |
| Observations | 532 | 532 |
| Interaction | No | Yes |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the probit model with an interaction between being married and being Muslim, the constant term is the only statistically significant coefficient, as in the original probit model. Interestingly, the magnitude of the coefficients for both the married and Muslim terms is much larger in magnitude than in the model without the interaction term, but with larger variance as well.

## Question 13

Using the margins command, I find a mean interact effect of -0.6719, and by hand I find a mean finite difference of -0.0657 with the interaction term.

## Question 14

I compute the variance of the finite differences for the interaction effect to be $8 \times 10^{-5}$ which is small due to the small coefficient estimates.

## Question 15

Table 8. Regression of the LPM Squared Residuals

| VARIABLES | (1)<br>residuals2 |
|---|---|
| client_age | 0.000168 |
| | (0.00142) |
| client_married | 0.00774 |
| | (0.0348) |
| client_education | -0.00213 |
| | (0.00271) |
| hh_size | -0.00786 |
| | (0.00612) |
| hh_income | 2.51e-06 |
| | (2.42e-06) |
| muslim | -0.00633 |
| | (0.0241) |
| hindu_sc_kat | -0.0168 |
| | (0.0346) |
| treated | 0.0282 |
| | (0.0228) |
| Constant | 0.150** |
| | (0.0752) |
| Observations | 532 |
| R-squared | 0.008 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Since the constant is the only statistically significant coefficient, heteroskedasticity does not seem to be a concern.

## Question 16

Table 9. Probit Model with Heteroskedastic Errors

| | (1) | (2) |
|---|---|---|
| VARIABLES | taken_new | lnsigma |
| client_age | -0.112 | 0.0285 |
| | (0.137) | (0.0196) |
| client_married | 0.129 | |
| | (0.846) | |
| client_education | -0.311 | 0.0694 |
| | (0.256) | (0.0474) |
| hh_size | -0.226 | |
| | (0.197) | |
| hh_income | 6.93e-05 | |
| | (7.23e-05) | |
| muslim | -0.179 | |
| | (0.584) | |
| hindu_sc_kat | -0.344 | |
| | (0.962) | |
| treated | 0.915 | |
| | (0.787) | |
| Constant | 1.715 | |
| | (3.573) | |
| Observations | 532 | 532 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In order to complete a likelihood ratio test of lnsigma=0, we will consider a Chi-squared distribution with 2 degrees of freedom and a critical value of 2.97. However, in this case the probability is 0.2267, so we cannot reject the null of homoskedasticity.