

# Econ 717: problem set 1

Emily Case

February 16, 2022

## Problem 1

## Problem 2

(Table 1, column (1)) There is a negative relationship between new loan take up and client age (though it's very small), client education, household size, Muslim, and Hindi. Though none of the coefficients are statistically significant, the constant term seems to have the highest impact on new loan take up, implying that unmarried, uneducated, young, non-Muslim and non-Hindi households that are small and poor are most likely to take up a new loan.

## Problem 3

Table 1 column (2) compares the results from problems 2 and 3. notice that the robust standard errors are larger for client age, household size, household income, and the constant. Although, it seems these differences are not huge, which might mean there is not much heteroskedasticity.

## Problem 4

The minimum value is about 0.058, and the maximum is 0.282; none of the predicted probabilities lie outside of  $[0,1]$ .

## Problem 5

Results are in Table 1, column (3). The coefficients are the same, and the standard error results are not that different from the results in (1) or (2).

Table 1: Problems 2, 3, & 5			
	(1) LPM	(2) LPM, robust SE	(3) vWLS
Client age	-0.0000 (0.0022)	-0.0000 (0.0023)	-0.0000 (0.0021)
Client marital status	0.0117 (0.053)	0.0117 (0.052)	0.0117 (0.053)
Client education	-0.0037 (0.0041)	-0.0037 (0.0041)	-0.0037 (0.0041)
Household size	-0.0113 (0.0093)	-0.0113 (0.0093)	-0.0113 (0.0093)
Household income	0.0000 (0.0000037)	0.0000 (0.0000037)	0.0000 (0.0000037)
Client is Muslim	-0.0076 (0.037)	-0.0076 (0.037)	-0.0076 (0.037)
Client is Hindi	-0.0275 (0.053)	-0.0275 (0.051)	-0.0275 (0.052)
Treated	0.0426 (0.035)	0.0426 (0.033)	0.0426 (0.035)
Constant	0.1994 (0.11)	0.1994 (0.12)	0.1994 (0.11)
Observations	532	532	532

Standard errors in parentheses

## Problem 6

Results are in Table 2, columns (3) and (4). They should not be the same as the LPM estimates because the coefficients that probit and logit spit out are not the conditional probability, as they are with LPM. They are not completely different, but there are clearly differences between probit and logit as well.

Table 2: Problem 6				
	(1)	(2)	(3)	(4)
	LPM	LPM, robust SE	Probit	Logit
main				
Client age	-0.0000 (0.0022)	-0.0000 (0.0023)	0.0002 (0.0086)	-0.0004 (0.016)
Client marital status	0.0117 (0.053)	0.0117 (0.052)	0.0495 (0.21)	0.0931 (0.39)
Client education	-0.0037 (0.0041)	-0.0037 (0.0041)	-0.0146 (0.017)	-0.0276 (0.030)
Household size	-0.0113 (0.0093)	-0.0113 (0.0093)	-0.0476 (0.038)	-0.0854 (0.069)
Household income	0.0000 (0.0000037)	0.0000 (0.0000037)	0.0000 (0.000014)	0.0000 (0.000025)
Client is Muslim	-0.0076 (0.037)	-0.0076 (0.037)	-0.0326 (0.15)	-0.0533 (0.26)
Client is Hindi	-0.0275 (0.053)	-0.0275 (0.051)	-0.1100 (0.22)	-0.2078 (0.39)
Treated	0.0426 (0.035)	0.0426 (0.033)	0.1751 (0.14)	0.3187 (0.26)
Constant	0.1994 (0.11)	0.1994 (0.12)	-0.8527 (0.46)	-1.3740 (0.84)
Observations	532	532	532	532

Standard errors in parentheses

## Problem 7

The derivatives mean the change in the conditional probability of new loan take up resulting from a small change in client age. Results are below. The partial derivatives in all cases are pretty similar to each other.

	(1) dydx_reg	(2) dydx_logit	(3) dydx_probit
main			
Client age	-0.0000283 (0.00216)	-0.0000525 (0.00215)	0.0000382 (0.00213)
Observations	532	532	532

Marginal effects; Standard errors in parentheses

(d) for discrete change of dummy variable from 0 to 1

Table 3: Problem 7

Probit model derivatives, 4 ways		
	method	derivative
(a)	dprobit	.0000382
(b)	using formula & summarize	.0000606
(c)	numerical derivatives	.0000587
(d)	margins	.0000382

## Problem 8

The LPM derivative is much higher than the probit one.<sup>1</sup>

Table 4: Problem 8: comparing numerical derivatives

Probit	LPM (with quartic client age)
.0000587	.0135601

## Problem 9

I get an LRI of 0.009, which is very low, suggesting that the model does not do a good job of explaining new loan take up.

## Problem 10

Since the maximum value of the predicted probabilities was lower than 0.5, it makes sense that using 0.5 as a cut off would yield a 0% correct prediction rate. When changing it to the loan take up fraction (about 0.16), the correct prediction rate is about 51%.

Table 5: Problem 10

Cutoff value	Correct prediction rate
0.5	0
loan take up fraction	.5132

---

<sup>1</sup>I am pretty sure I did this wrong.

## Problem 11

The result of the same method, but with a subsample of the dataset, is not too different, which is what we would want to see.

Table 6: Problem 11

Cutoff value	Correct prediction rate	Correct prediction rate (subsample)
0.5	0	0
loan take up fraction	.5132	.4887

## Problem 12

Table 7: Problem 12: probit models

	(1) no interaction	(2) interaction
Client took a loan		
Client age	0.0002 (0.0086)	0.0009 (0.0087)
Client marital status=1	0.0495 (0.2141)	0.1557 (0.2807)
Client education	-0.0146 (0.0166)	-0.0154 (0.0167)
Household size	-0.0476 (0.0379)	-0.0502 (0.0381)
Household income	0.0000 (0.0000)	0.0000 (0.0000)
Client is Muslim=1	-0.0326 (0.1468)	0.2059 (0.4206)
Client is Hindi	-0.1100 (0.2152)	-0.1143 (0.2153)
Treated	0.1751 (0.1419)	0.1820 (0.1424)
Client marital status=1 × Client is Muslim=1		-0.2709 (0.4481)
Constant	-0.8527 (0.4592)	-0.9589 (0.4939)
Observations	532	532

Standard errors in parentheses

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

## Problem 13

Table 8: Problem 13

		standard error
interaction effects	-.016	.0381
finite differences	-.0644	

## Problem 14

I include the standard error in Table 8 from problem 13.

## Problem 15

We are interested if the covariates help explain the variation in the estimated squared residuals. In Table 9, since the  $R^2$  is the same for both LPM and the model with squared residuals, we do not see evidence for heteroskedasticity.

## Problem 16

Since the coefficients in column (4) of Table 9 are very different from the probit model before allowing for heteroskedasticity, we can infer that there is evidence for heteroskedasticity.<sup>2</sup>

---

<sup>2</sup>I did not really understand how to read the results for problems 15 or 16.



Table 9: Problems 15 &amp; 16

	(1) LPM	(2) Squared residuals	(3) Probit	(4) hetprobit
main				
Client age	-0.0000 (0.0023)	0.0002 (0.0015)	0.0002 (0.0086)	-0.1123 (0.1375)
Client marital status	0.0117 (0.0519)	0.0077 (0.0350)	0.0495 (0.2141)	0.1288 (0.8459)
Client education	-0.0037 (0.0041)	-0.0021 (0.0027)	-0.0146 (0.0166)	-0.3106 (0.2555)
Household size	-0.0113 (0.0093)	-0.0079 (0.0063)	-0.0476 (0.0379)	-0.2261 (0.1970)
Household income	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0001 (0.0001)
Client is Muslim	-0.0076 (0.0365)	-0.0063 (0.0240)	-0.0326 (0.1468)	-0.1792 (0.5843)
Client is Hindi	-0.0275 (0.0510)	-0.0168 (0.0356)	-0.1100 (0.2152)	-0.3443 (0.9619)
Treated	0.0426 (0.0335)	0.0282 (0.0230)	0.1751 (0.1419)	0.9150 (0.7867)
Constant	0.1994 (0.1169)	0.1505 (0.0771)	-0.8527 (0.4592)	1.7151 (3.5729)
Insigma				
Client age				0.0285 (0.0196)
Client education				0.0694 (0.0474)
Observations	532	532	532	532
$R^2$	0.008	0.008		

Standard errors in parentheses

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