717 Metrics

Giselle Labrador Badia

February 2022

This assignment uses data from Field et al. (2010). Tables with standards errors are provided for all regressions and other relevant analyses that I discuss. All the Stata code pertinent to this assignment is attached.

Question 1. Missing values of the suggested covariates were dropped. Also, some observations with age 0 were dropped since it appears that those are not properly coded.

VARIABLES	(1) LPM	(2) RLPM	(3) WLS	(4) Probit	(5) Logit	(6) QLPM
Treated	0.0431	0.0431	0.0431	0.176	0.320	0.0470
	(0.0341)	(0.0327)	(0.0340)	(0.139)	(0.253)	(0.0338)
Client_Age	-0.000377	-0.000377	-0.000377	-0.00107	-0.00295	-0.502
	(0.00210)	(0.00222)	(0.00209)	(0.00829)	(0.0152)	(0.176)
Client_Married	0.0186	0.0186	0.0186	0.0814	0.147	0.0236
	(0.0500)	(0.0481)	(0.0499)	(0.203)	(0.368)	(0.0524)
Client_Education	-0.00510	-0.00510	-0.00510	-0.0201	-0.0376	-0.00475
	(0.00397)	(0.00401)	(0.00396)	(0.0159)	(0.0287)	(0.00395)
HH_Size	-0.0109	-0.0109	-0.0109	-0.0462	-0.0821	-0.00862
	(0.00918)	(0.00910)	(0.00915)	(0.0372)	(0.0679)	(0.00916)
HH_Income	4.04e-06	4.04e-06	4.04e-06	1.66e-05	2.83e-05	4.55e-06
	(3.64e-06)	(3.72e-06)	(3.63e-06)	(1.41e-05)	(2.47e-05)	(3.63e-06)
muslim	-0.0171	-0.0171	-0.0171	-0.0691	-0.122	-0.0226
	(0.0361)	(0.0358)	(0.0361)	(0.145)	(0.260)	(0.0359)
Hindu_SC_Kat	-0.0248	-0.0248	-0.0248	-0.0995	-0.186	-0.0299
	(0.0512)	(0.0499)	(0.0510)	(0.207)	(0.377)	(0.0508)
age2						0.0202
						(0.00678)
age3						-0.000341
						(0.000110)
age4						2.04e-06
						(6.31e-07)
Constant	0.210	0.210	0.210	-0.819	-1.301	4.615
	(0.111)	(0.112)	(0.110)	(0.444)	(0.807)	(1.631)

Standard errors in parentheses

Table 1: Regressions result for questions 1-6. It also includes the quartic regression for question 8 (6).

Question 2. Column 1 of Table 1 presents the regressions result for the LPM, which tries to explain loan take-up. None of the coefficients of the covariates have statistical significance and in particular, the treatment variable. There are many issues with this model, so this is not surprising. For instance, the dependent variable it is not treated as a binary variable. Also, age and years of education are treated as continuous, where brackets could do a better job in finding the effect of loan take-up.

Question 3. Column 1 of Table 1 presents the regression results for the LPM with robust standard errors. Again, none of the coefficients of the covariates have statistical significance and these coefficients coincide with the LPM regression coefficients. Standard errors are for most of the covariates slightly smaller, which suggests that there is evidence against heteroskedasticity.

Question 4. Below, I display the mean standard deviation, minimum and maximum values of the LPM probability predictions of loan take-up. Probabilities all lie between the [0,1] interval. Not only that, but the maximum value is around 0.303, which is smaller than 0.5. This indicates that individual take-up is overall not very likely. Also, we should be careful when using a cutoff of 0.5 and calculating correct prediction rates.

	Mean	SD	Min	Max	N
Linear prediction	.1693694	.0373802	.0342446	.3029885	555.00

Question 5. Column 3 of Table 1 displays the estimation of the model by weighted least squares using Stata's *vwls*. The coefficients do not differ from those of the previous regressions. The standard errors are in some cases smaller than LPM and robust LPM (Client_Age, Client_Education) and other cases comparable with LPM (Client_Married, muslin, Treated). Overall, some standard errors are smaller than the robust specification and some are larger.

Question 6. Columns 4 and 5 present regression results for the probit and logit models. As expected, Probit and Logit coefficients are very different than those of the LPM and WLS; coefficients are larger for all independent variables. For the treatment variable, logit's coefficient is greater than probit's coefficient, although these are still non-significant. These coefficients should be different since probit and logit coefficients indicate the constant marginal effect of a variable on the underlying latent variable. Meanwhile, the LPM coefficients encapsulate this effect, but on the dependent variable probability.

Question 7. I show in Table 2 the different Marginal Partial Effects for LPM, Logit, and Probit. LPM and logit effects differ from each other and probit MPE. Moreover, probit coefficients are very similar across the 4 different methods. For logit and probit, MPE means the average effect of one year increase in age on the latent scoring take-up variable or the average increase in probability. For LPM, MPE represents the average effect of one year increase in age on the probability of loan take-up.

	LPM	Logit	Probit	QLPM
MPE	-.0003774	-.0004105	-	-
dprobit	-	-	-.0002683	-
by hand	-	-	-.0002681	-
small variation	-	-	-.0002682	-.0024352
margins	-	-	-.0002681	-

Table 2: Marginal Partial Effects for different specifications and different method for questions 7 and 8.

Question 8. The quartic LPM, shows MPE closer to those of probit. This effect is shown in Table 2, column 4. This suggests that with greater flexibility the LPM does better.

Question 9. The LRI obtained by hand is 0.073. This index is always between 0 and 1, hence, we interpret that the model does not do a good job in explaining the variance.

Question 10. Table 3 shows the predictive success by estimating the probit model using both a 0.5 cutoff and the sample fraction. The sample loan take-up rate obtained is 0.168. Between the two measures of goodness of fit, I prefer the second one. The cutoff of 0.5, yields a prediction rate equal to the number of 0s or no loan take-ups. This is because most people do not take up loans, and the prediction probabilities are for all observations under 0.5. This prediction rate is therefore not very informative, and it will make the econometrician think that the model is better than it is. Similar models will all give the same prediction rate, this is why I prefer the second measure of goodness of fit.

| | In-sample | | Out-of-sample | |
	cutoff 0.5	take-up rate	cutoff 0.5	take-up rate
Probit	.831	.541	.826	.562

Table 3: Correct probability rates in-sample and out-of-sample using two methods. (Question 10 and 11)

Question 11. These results are also in Table 3. In and out-of-sample give similar results for the two predictive performance methods. Actually, out-of-sample indicates a slightly better prediction rate for the mean probability than in-sample. Out-of-sample is preferred, and again, take-up rates are more appropriate to assess take-up since is far from being equiprobable.

Question 12. Column 1 of Table 4 shows the regression results with an interaction term between Muslim and married. The standard errors are slightly smaller but the statistical significance and coefficient values do not change much.

Question 13. The table below shows the mean finite difference for Muslim, married, and both, in the models with and without interaction. This marginal effect suggests that married Muslims are less likely to take up a loan than unmarried Muslims and not Muslims who are married. The effects without the interaction term make less sense.

| | | Mean Finite Difference | |
married	muslim	not muslin and married	muslim and married
1	0	-.02754	.04446
0	1	.02399	.04421
1	1	-.00355	.01810

Table 4: Mean finite differences (question 13)

Question 14. The standard deviation of the estimated interaction effects obtained in Problem 13 across sample observations is .01058. There is a large percent of observations of men that are both Muslim and married. This might help explain why the variance is so small.

Question 15. Table 6 presents the results of the regression of the squared residuals of the LPM on the independent variables. Almost no coefficient is statistically significant. This may be due to heteroskedasticity in the client age variable, which is the only significant coefficient. However, I run an

heteroskedasticity test, and we fail to reject the null (no heteroskedasticity), therefore I find no evidence of heteroskedasticity.

Question 16. Columns 2 and 3 of Table 5 display the results of the regression using the *hetprob* command. From the result of the LR test (last line of the table), I conclude there is no evidence of heteroskedasticity.

VARIABLES	(1) muslim*Client_Married	(2) Hetprob	(3) lnsigmas (Hetprob)
Treated	0.184	1.011	
	(0.139)	(0.777)	
Client_Age	-0.000273	-0.118	0.0283
	(0.00840)	(0.119)	(0.0168)
Client_Married	0.187	0.224	
	(0.259)	(0.831)	
Client_Education	-0.0209	-0.380	0.0800
	(0.0160)	(0.258)	(0.0437)
HH_Size	-0.0492	-0.232	
	(0.0375)	(0.187)	
HH_Income	1.61e-05	9.54e-05	
	(1.42e-05)	(7.80e-05)	
muslim	0.188	-0.351	
	(0.403)	(0.597)	
Hindu_SC_Kat	-0.105	-0.198	
	(0.207)	(0.848)	
1.mus_married	-0.294		
	(0.431)		
Constant	-0.923	1.791	
	(0.471)	(3.140)	

LR test of lnsigma=0: chi2(2) = 4.48   Prob > chi2 = 0.1062

Standard errors in parentheses

Table 5: Regression results for question 12 and 16.

	(1)
VARIABLES	Resid2
Client_Age	3.97e-05
	(0.00137)
Client_Married	0.0138
	(0.0326)
Client_Education	-0.00291
	(0.00259)
HH_Size	-0.00778
	(0.00598)
HH_Income	3.05e-06
	(2.37e-06)
muslim	-0.0117
	(0.0236)
Hindu_SC_Kat	-0.0156
	(0.0334)
Treated	0.0281
	(0.0222)
Constant	0.153
	(0.0721)

Standard errors in parentheses

Table 6: Regression results for squared residuals regression.