ECON 717A: Problem Set 1

Alex von Hafften

February 12, 2022

1 Write-Up

Problem 1 - Dropping Miss Values

I drop observations with missing values for `HH_Income` as well as observations with indicators `miss_Client_Age`, `miss_Client_Married`, and `miss_Client_Education` equaling one. This filtering drops 65 observations.

Problem 2 - LPM

I estimate a linear probability model with homoskedastic standard errors of `taken_new` on `Client_Age`, `Client_Married`, `Client_Education`, `HH_Size`, `HH_Income`, `muslim`, `Hindu_SC_Kat`, and `Treated`.

	(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

The difference between all coefficients and zero is statistically insignificant even at the 10 percent level; the constant is statistically significant at the 10 percent level. Consistent with the insignificant coefficients, the constant estimate is about 0.2, which roughly corresponds to the unconditional probability of taking up a new loan of about 17 percent.

Problem 3 - LPM with Robust SE

I estimated a linear probability model with comma robust standard errors of `taken_new` on the same set of covariates. In the first column, I report homoskedastic standard errors (exactly the same as the table in problem 1) and, in the second column, I report comma robust standard errors.

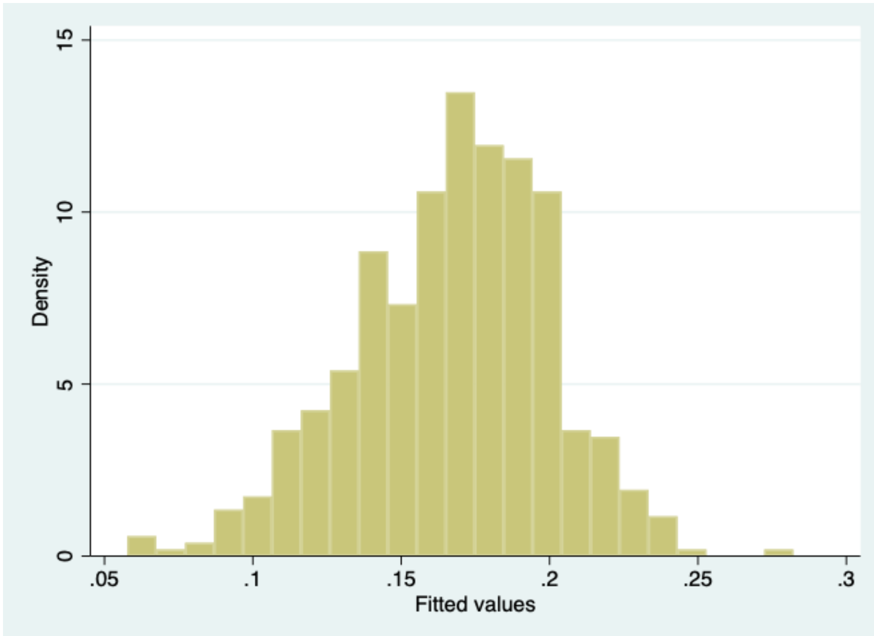
VARIABLES	(1) taken_new	(2) taken_new
Client_Age	-2.83e-05	-2.83e-05
	(0.00216)	(0.00227)
Client_Married	0.0117	0.0117
	(0.0529)	(0.0519)
Client_Education	-0.00369	-0.00369
	(0.00412)	(0.00410)
HH_Size	-0.0113	-0.0113
	(0.00931)	(0.00928)
HH_Income	3.14e-06	3.14e-06
	(3.68e-06)	(3.71e-06)
muslim	-0.00756	-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
Comma Robust SEs	No	Yes

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

Compared to the homoskedastic standard errors, the comma robust standard errors are larger only for `Client_Age`, `HH_Income`, and the constant. For all other covariates, the comma robust standard errors are smaller. Despite the smaller standard errors for these variables, the constant coefficient is still the only statistically significant coefficient (albeit at the 10 percent level).

Problem 4 - Predicted Probability

I predict the probabilities from the LPM and plot a histogram:



As seen in the histogram, all predicted probabilities are between zero and one.

Problem 5 - Weighted Least Squares

I tried running the baseline specification with using variance weighted least squares, but I get the following error: *no groups with sufficient observations*. I suspect this error occurs because we do not have enough observation to compute the by-group variances to weight observations. I am able to estimate a more parsimonious model using variance weighted least squares model by dropping religious/caste indicators, `muslim` and `Hindu_SC_Kat`, as well as `Client_Married` due to multicollinearity. Column (1) is the more parsimonious specification estimated using unweighted least squares and column (2) is that estimated using variance weighted least squares.

	(1)	(2)
VARIABLES	taken_new	taken_new
Client_Age	-4.51e-05	-0.000535
	(0.00224)	(0.0198)
Client_Education	-0.00325	0.00214
	(0.00400)	(0.0184)
HH_Size	-0.0102	0.00353
	(0.00936)	(0.0576)
Treated	0.0412	-0.125
	(0.0332)	(0.373)
Constant	0.216**	0.490
	(0.107)	(0.877)
Observations	532	65
R-squared	0.006	
Weighted By	None	Variance

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Coefficients in both regressions above are insignificant, which is consistent with the less parsimonious unweighted baseline linear probability estimates above. Without weighting, the coefficients on `Client_Education` and `HH_Size` are larger compared to their variance weighted counterparts while the other coefficients are smaller.

Problem 6 - Probit and Logit

In the table below, there are estimates for the LPM with comma robust SEs (column 1), probit (column 2), and logit (column 3).

	(1)	(2)	(3)
VARIABLES	taken_new	taken_new	taken_new
Client_Age	-2.83e-05	0.000154	-0.000382
	(0.00227)	(0.00856)	(0.0157)
Client_Married	0.0117	0.0495	0.0931
	(0.0519)	(0.214)	(0.388)
Client_Education	-0.00369	-0.0146	-0.0276
	(0.00410)	(0.0166)	(0.0300)
HH_Size	-0.0113	-0.0476	-0.0854
	(0.00928)	(0.0379)	(0.0694)
HH_Income	3.14e-06	1.33e-05	2.23e-05
	(3.71e-06)	(1.44e-05)	(2.53e-05)
muslim	-0.00756	-0.0326	-0.0533
	(0.0365)	(0.147)	(0.263)
Hindu_SC_Kat	-0.0275	-0.110	-0.208
	(0.0510)	(0.215)	(0.395)
Treated	0.0426	0.175	0.319
	(0.0335)	(0.142)	(0.259)
Constant	0.199*	-0.853*	-1.374
	(0.117)	(0.459)	(0.835)
Observations	532	532	532
R-squared	0.008		
Model	LPM	Probit	Logit

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The probit and the logit coefficients are not the same but they are similar, especially when comparing them to the LPM coefficients. These results make sense. The probit uses the standard normal cdf as its link function, the logit uses its logistic function as the link function, and the LPM uses $f(x) = x$ as its link function. The standard normal cdf and the logistic function are much closer to each than they are to $f(x) = x$.

Problem 7 - Mean Partial Derivatives

I compute the mean partial derivative for `Client_Age` using six methods:

1. LPM coefficient.
2. `dprobit`
3. Analytically using a probit model. The mean partial derivative estimated as $\phi(x\beta')\beta_1$ where ϕ is the pdf of the standard normal.
4. Numerically using a probit model. Estimate the probit model and predict the probability: $\hat{p}_i$. Perturb `Client_Age` by $\varepsilon = 0.01$ and reestimate the probit model to predict the probability: $\hat{p}_i^{\varepsilon}$. Compute the partial derivative for each observations: $\frac{\hat{p}_i - \hat{p}_i^{\varepsilon}}{\varepsilon}$. Compute the average across observations.
5. `Margins` for probit.
6. `Margins` for logit.

Model	Approach	Mean Partial Derivative Estimate
LPM	-	-0.0000283
Probit	`dprobit`	0.0000382
Probit	Analytical	0.0000382
Probit	Numerical	0.0000386
Probit	`margins`	0.0000382
Logit	`margins`	-0.0000525

All the estimates are very close to zero, but some are positive and some are negative. The differences between the mean partial derivative estimates partially boil down to how noisy the estimate of the coefficients on `Client_Age` is. In every regression (LPM, probit, and logit), the coefficient is statistically indistinguishable from zero. Thus, the differences in the mean partial derivative estimates boil out to noise. For the probit-based estimates, all but the numerical estimates are the same. This must be due to `dprobit` and `margins` under the analytical derivative "under the hood."

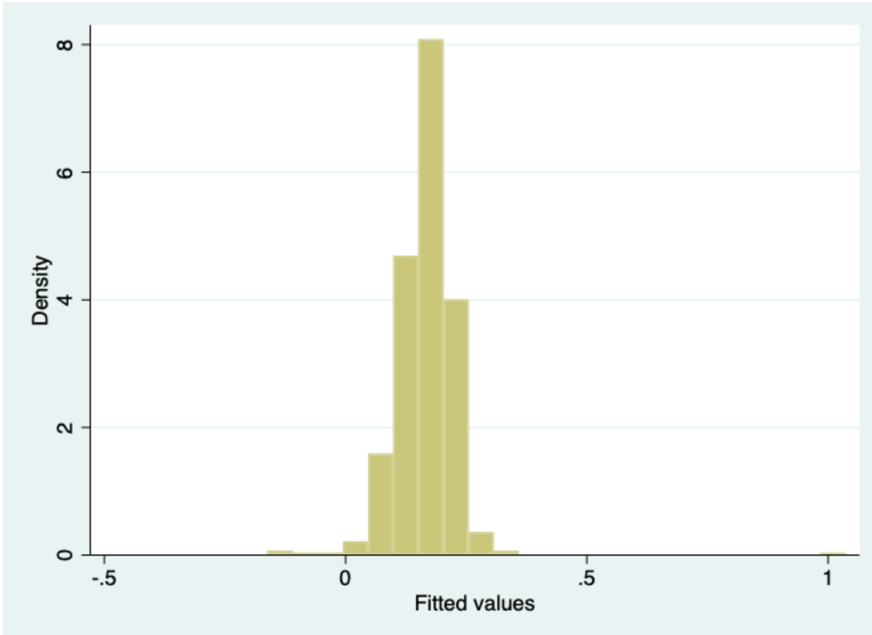
Problem 8 - LPM with Quartic Age

I add quadratic transformations of `Client_Age` to the regression. Column (1) is the baseline model and column (2) includes the additional terms.

VARIABLES	(1) taken_new	(2) taken_new
Client_Age		-0.493***
		(0.168)
Client_Age_2		0.0199***
		(0.00621)
Client_Age_3		-0.000336***
		(9.68e-05)
Client_Age_4		2.01e-06***
		(5.35e-07)
Client_Married	0.0117	0.0152
	(0.0518)	(0.0552)
Client_Education	-0.00368	-0.00315
	(0.00394)	(0.00411)
HH_Size	-0.0113	-0.00889
	(0.00927)	(0.00924)
HH_Income	3.14e-06	3.65e-06
	(3.71e-06)	(3.70e-06)
muslim	-0.00755	-0.0128
	(0.0363)	(0.0361)
Hindu_SC_Kat	-0.0275	-0.0333
	(0.0510)	(0.0507)
Treated	0.0426	0.0465
	(0.0335)	(0.0334)
Constant	0.198**	4.518***
	(0.0779)	(1.605)
Observations	532	532
R-squared	0.008	0.031

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

By their statistical significance, it appears that it is important to include these higher-order polynomial transformations of `Client_Age`. I numerically computing the mean partial derivative with respect to `Client_Age` and get 2.32e-07. So yes, adding the quadratic terms “does better” in so much as the estimate is close to the probit model estimates. However, there are now six observations with predicted probabilities outside the unit interval (five less than zero and one greater than one).



Problem 9 - LRI

The log-likelihood of the baseline probit model is $\ln \hat{L} = -238.0747$ and the log-likelihood of a probit model with only a constant is $\ln L_0 = -240.2343$. Thus, the LRI is $LRI = 1 - \frac{\ln \hat{L}}{\ln L_0} = 0.0090$. This matches the Stata output for pseudo-R^2. Since the LRI is low, this would suggest that the additional covariates do not explain more variation in outcome beyond what is captured by the constant.

Problem 10 - Correction Prediction Rates

First, I consider correct prediction rates based on a 50 percent cutoff. The idea behind the 50 percent cutoff is whether you're more likely to take out a loan than not. I get the following table:

	$\hat{p} < 0.5$	$\hat{p} > 0.5$
`taken_new` $= 1$	0.8327	0.0
`taken_new` $= 0$	0.1673	0.0

This table is due to having none of the predicted probabilities are over 50 percent. The correct prediction rate here is 0.41635 (i.e., 0.8327 times 50 percent). Second, I consider correct prediction rates based on a cutoff equal to unconditional probability of the outcome, 0.1673. The idea behind this cutoff is whether you're more likely than a randomly selected person from the sample to take out a loan. I get the following table:

	$\hat{p} < 0.1673$	$\hat{p} > 0.1673$
`taken_new` $= 1$	0.4267	0.4060
`taken_new` $= 0$	0.0602	0.1071

The correct prediction rate is now 0.2669 (i.e., 0.4267 times 50 percent plus 0.1071 times 50 percent).

Problem 11 - In-Sample vs. Out-of-Sample Correct Prediction Rates

My expectation is that in-sample correct prediction rates will be higher than out-of-sample correct prediction rates. First, based on a 50 percent cutoff, I get the following table for the estimation subsample:

	$\hat{p} < 0.5$	$\hat{p} > 0.5$
`taken_new` = 1	0.8383	0.0
`taken_new` = 0	0.1617	0.0

Thus, the correct prediction rate is 0.4191. I get the following table for the non-estimation subsample:

	$\hat{p} < 0.5$	$\hat{p} > 0.5$
`taken_new` = 1	0.8271	0.0
`taken_new` = 0	0.1729	0.0

Thus, the correct prediction rate is 0.4135. The higher correct prediction rate for the estimation sample has nothing to do with the accuracy of in-sample vs. out-of-sample prediction because all predicted values are less than 50 percent. The difference is only due to sample variation. Second, for a cutoff equal to unconditional probability of the outcome (0.1673), I get the following table for the estimation subsample:

	$\hat{p} < 0.1673$	$\hat{p} > 0.1673$
`taken_new` = 1	0.4812	0.3571
`taken_new` = 0	0.0827	0.0789

Thus, the correct prediction rate is 0.2801. I get the following table for the non-estimation subsample:

	$\hat{p} < 0.1673$	$\hat{p} > 0.1673$
`taken_new` = 1	0.4436	0.3835
`taken_new` = 0	0.1128	0.0602

Thus, the correct prediction rate is 0.2519. This difference may be a confirmation that the prediction rate in-sample is better than out-of-sample.

Problem 12 - Interaction Term

Below is the baseline probit model and the probit model with an interaction term for married and Muslim.

	(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

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

Problem 13 - Interaction Term Finite Differences

I compute the interaction effect both with and without the terms highlighted by Ai and Norton (2003). For the interaction effect without these terms, I use `margin` which automatically computes finite differences given a binary independent variable. The estimate of this interaction effect is -0.06719. To compute the interaction effect with the terms highlighted by Ai and Norton (2003), I compute the predicted index from the probit. Then I subtract the coefficient for each three dummies (married, Muslim, and married $\times$ Muslim). This index, I_0, corresponds to the predicted value conditional on being unmarried and not Muslim for all observations. Then I create a three variables based on this index: I_1 adds in the coefficient on married, I_2 adds in the coefficient on Muslim, and I_{12} adds in all three coefficients. Then the interaction affect for each observation is $\Phi(I_{12}) - \Phi(I_1) - \Phi(I_2) + \Phi(I_0)$. The average of these finite differences is -0.0657. So the Ai and Norton (2003) terms slightly attenuate the estimated interaction effect.

Problem 14 - Interaction Term Finite Differences Variance

I compute the variance of the finite differences for the interaction effect to be 0.00008. The small variance stems from all coefficient estimating being quite small.

Problem 15 - Heteroskedasticity Test

I compute the residuals from the baseline LPM, square them, and regress them on the usual covariates:

VARIABLES	(1) residuals_p_2
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

All coefficients are insignificant except for the constant. These results indicate that heteroskedasticity is not a concern.

Problem 16 - Probit with Heteroskedasticity

The probit model with heteroskedasticity of the error term as a function of `Client_Age` and `Client_Education` is below.

	(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

The critical value for the likelihood ratio test of `lnsigma=0` is 2.97 and it is distributed chi-squared with 2 degrees of freedom. The resulting p-value is 0.2267, so we fail to reject the null of homoskedasticity. This result confirms the finding in problem 15 of a lack of heteroskedasticity.

2 Stata Log File

```
------------------------------------------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/alexandervonhafften/Documents/UW Madison/problem_sets/econ_717a/ps1/analysis.smcl
  log type:  smcl
 opened on:  12 Feb 2022, 13:29:44

.
. * Alex von Hafften

. * Problem set 1

. * ECON 717A: Applied Economics

.
. * clear workspace

. clear

.
. * install user defined functions (if needed)

. ssc install outreg2
checking outreg2 consistency and verifying not already installed...
all files already exist and are up to date.

.
. * change working directory

. cd "/Users/alexandervonhafften/Documents/UW Madison/problem_sets/econ_717a/ps1/"
/Users/alexandervonhafften/Documents/UW Madison/problem_sets/econ_717a/ps1

.
. * open dataset

. use "Field et al (2010) Analysis Sample"
(ATTENTION: Type notes in the command line for information on this dataset)
```

```
.
. ******************************************************************************

. * problem #1 - drop missing values

. ******************************************************************************

.
. drop if missing(HH_Income)
(36 observations deleted)

. drop if miss_Client_Age == 1
(6 observations deleted)

. drop if miss_Client_Married == 1
(7 observations deleted)

. drop if miss_Client_Education == 1
(16 observations deleted)

.
. ******************************************************************************

. * problem #2 - estimate linear probability model with homoskedastic standard errors

. ******************************************************************************

.
. * define list of variable in usual_covariate

. local covariates " Client_Married Client_Education HH_Size HH_Income muslim Hindu_SC_Kat Treated"

.
. * estimate LPM

. regress taken_new Client_Age `covariates'

      Source |       SS           df       MS      Number of obs   =       532
-------------+----------------------------------   F(8, 523)       =      0.52
```

```
       Model |  .586701556         8  .073337695   Prob > F        =    0.8405
    Residual |  73.5242007       523  .140581646   R-squared       =    0.0079
-------------+----------------------------------   Adj R-squared   =   -0.0073
       Total |  74.1109023       531  .139568554   Root MSE        =    .37494

------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |  -.0000283   .0021572    -0.01   0.990     -.004266    .0042095
  Client_Married |   .0117141   .0529309     0.22   0.825    -.0922693    .1156975
Client_Education |  -.0036931   .0041228    -0.90   0.371    -.0117924    .0044062
         HH_Size |  -.0113264   .0093055    -1.22   0.224    -.0296072    .0069544
       HH_Income |   3.14e-06   3.68e-06     0.85   0.393    -4.08e-06    .0000104
          muslim |  -.0075629   .0366854    -0.21   0.837    -.0796318    .0645061
    Hindu_SC_Kat |  -.0274764   .0526227    -0.52   0.602    -.1308542    .0759015
         Treated |   .0426031   .0346907     1.23   0.220    -.0255471    .1107532
           _cons |   .1993902   .1142909     1.74   0.082    -.0251353    .4239158
------------------------------------------------------------------------------------

. outreg2 using p2_table, tex(frag) replace
p2_table.tex
dir : seeout

.
. ********************************************************************************

. * problem #3 - estimate linear probability model with heteroskedastic standard errors

. ********************************************************************************

.
. regress taken_new Client_Age ‘covariates’

      Source |       SS           df       MS      Number of obs   =       532
-------------+----------------------------------   F(8, 523)       =      0.52
       Model |  .586701556         8  .073337695   Prob > F        =    0.8405
    Residual |  73.5242007       523  .140581646   R-squared       =    0.0079
-------------+----------------------------------   Adj R-squared   =   -0.0073
       Total |  74.1109023       531  .139568554   Root MSE        =    .37494
```

```
------------------------------------------------------------------------------
   taken_new | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  Client_Age |  -.0000283   .0021572    -0.01   0.990     -.004266    .0042095
Client_Married |   .0117141   .0529309     0.22   0.825    -.0922693    .1156975
Client_Education |  -.0036931   .0041228    -0.90   0.371    -.0117924    .0044062
     HH_Size |  -.0113264   .0093055    -1.22   0.224    -.0296072    .0069544
   HH_Income |   3.14e-06   3.68e-06     0.85   0.393    -4.08e-06    .0000104
      muslim |  -.0075629   .0366854    -0.21   0.837    -.0796318    .0645061
 Hindu_SC_Kat |  -.0274764   .0526227    -0.52   0.602    -.1308542    .0759015
     Treated |   .0426031   .0346907     1.23   0.220    -.0255471    .1107532
       _cons |   .1993902   .1142909     1.74   0.082    -.0251353    .4239158
------------------------------------------------------------------------------

. outreg2 using p3_table, tex(frag) replace addtext(Comma Robust SEs, No)
p3_table.tex
dir : seeout

. regress taken_new Client_Age `covariates', robust

Linear regression                               Number of obs     =        532
                                                F(8, 523)         =       0.56
                                                Prob > F          =     0.8074
                                                R-squared         =     0.0079
                                                Root MSE          =     .37494

------------------------------------------------------------------------------
             |               Robust
   taken_new | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  Client_Age |  -.0000283   .0022749    -0.01   0.990    -.0044974    .0044409
Client_Married |   .0117141   .0518841     0.23   0.821    -.0902127    .1136409
Client_Education |  -.0036931   .0041013    -0.90   0.368    -.0117502    .0043641
     HH_Size |  -.0113264   .0092822    -1.22   0.223    -.0295613    .0069085
   HH_Income |   3.14e-06   3.71e-06     0.85   0.398    -4.15e-06    .0000104
      muslim |  -.0075629   .0365483    -0.21   0.836    -.0793625    .0642367
 Hindu_SC_Kat |  -.0274764   .0509678    -0.54   0.590    -.1276031    .0726504
     Treated |   .0426031   .0334686     1.27   0.204    -.0231463    .1083524
```

```
           _cons |   .1993902   .1168783     1.71   0.089    -.0302184    .4289989
----------------------------------------------------------------------------------

. outreg2 using p3_table, tex(frag) append addtext(Comma Robust SEs, Yes)
p3_table.tex
dir : seeout

.
. ********************************************************************************

. * problem #4 - predicted probabilities

. ********************************************************************************

.
. predict taken_new_hat_lpm
(option xb assumed; fitted values)

. histogram taken_new_hat_lpm
(bin=23, start=.05766389, width=.00975825)

. graph export p4_figure.png, replace
file /Users/alexandervonhafften/Documents/UW Madison/problem_sets/econ_717a/ps1/p4_figure.png saved as PNG format

. count if taken_new_hat_lpm > 1
  0

. count if taken_new_hat_lpm < 0
  0

.
. ********************************************************************************

. * problem #5 - weighted least squares

. ********************************************************************************

.
. vwls taken_new `covariates'
```

```
note: Client_Married omitted because of collinearity
note: Hindu_SC_Kat omitted because of collinearity

Variance-weighted least-squares regression      Number of obs     =         12
Goodness-of-fit chi2(0)    =       .            Model chi2(5)     =       0.00
Prob > chi2                =       .            Prob > chi2       =     1.0000
------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------
  Client_Married |          0  (omitted)
Client_Education |          0   .1234568     0.00   1.000    -.2419709    .2419709
         HH_Size |          0   .8894888     0.00   1.000    -1.743366    1.743366
       HH_Income |          0   .0010218     0.00   1.000    -.0020027    .0020027
          muslim |          0   1.007025     0.00   1.000    -1.973732    1.973732
    Hindu_SC_Kat |          0  (omitted)
         Treated |          0   1.302893     0.00   1.000    -2.553624    2.553624
           _cons |         .5   9.101962     0.05   0.956    -17.33952    18.33952
------------------------------------------------------------------------------

.
. * Doesn't work due to small sample, but works for more parsimonious model

. regress taken_new Client_Age Client_Education HH_Size  Treated, robust

Linear regression                               Number of obs     =        532
                                                F(4, 527)         =       0.73
                                                Prob > F          =     0.5703
                                                R-squared         =     0.0058
                                                Root MSE          =     .37391

------------------------------------------------------------------------------
                 |               Robust
       taken_new | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |  -.0000451   .0022433    -0.02   0.984     -.004452    .0043619
Client_Education |    -.00325   .0039959    -0.81   0.416    -.0110998    .0045997
         HH_Size |  -.0101567   .0093619    -1.08   0.278     -.028548    .0082346
         Treated |   .0412283   .0332486     1.24   0.216    -.0240878    .1065443
           _cons |    .215995    .106784     2.02   0.044     .0062205    .4257696
```

```
------------------------------------------------------------------------------------

. outreg2 using p5_table, tex(frag) addtext(Weighted By, None) replace
p5_table.tex
dir : seeout

. vwls taken_new Client_Age Client_Education HH_Size  Treated

Variance-weighted least-squares regression      Number of obs     =         65
Goodness-of-fit chi2(19)   =    4.20            Model chi2(4)     =       0.14
Prob > chi2                =  0.9998            Prob > chi2       =     0.9978
------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |  -.0005351   .0197968    -0.03   0.978    -.0393361     .038266
Client_Education |   .0021375   .0183518     0.12   0.907    -.0338314    .0381063
         HH_Size |   .0035293   .0576289     0.06   0.951    -.1094213    .1164798
         Treated |  -.1251191   .3727442    -0.34   0.737    -.8556843    .6054462
           _cons |   .4895395   .8772291     0.56   0.577    -1.229798    2.208877
------------------------------------------------------------------------------------

. outreg2 using p5_table, tex(frag) addtext(Weighted By, Variance) append
p5_table.tex
dir : seeout

.
. ********************************************************************************

. * problem #6 - probit and logit

. ********************************************************************************

.
. regress taken_new Client_Age ‘covariates’, robust

Linear regression                               Number of obs     =        532
                                                F(8, 523)         =       0.56
                                                Prob > F          =     0.8074
                                                R-squared         =     0.0079
```

```
                                                Root MSE          =     .37494

------------------------------------------------------------------------------
                 |               Robust
       taken_new | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |  -.0000283   .0022749    -0.01   0.990    -.0044974    .0044409
  Client_Married |   .0117141   .0518841     0.23   0.821    -.0902127    .1136409
Client_Education |  -.0036931   .0041013    -0.90   0.368    -.0117502    .0043641
         HH_Size |  -.0113264   .0092822    -1.22   0.223    -.0295613    .0069085
       HH_Income |   3.14e-06   3.71e-06     0.85   0.398    -4.15e-06    .0000104
          muslim |  -.0075629   .0365483    -0.21   0.836    -.0793625    .0642367
    Hindu_SC_Kat |  -.0274764   .0509678    -0.54   0.590    -.1276031    .0726504
         Treated |   .0426031   .0334686     1.27   0.204    -.0231463    .1083524
           _cons |   .1993902   .1168783     1.71   0.089    -.0302184    .4289989
------------------------------------------------------------------------------

. outreg2 using p6_table, tex(frag) replace addtext(Model, LPM)
p6_table.tex
dir : seeout

. probit taken_new Client_Age `covariates'

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
Iteration 3:   log likelihood = -238.07469

Probit regression                                       Number of obs =    532
                                                        LR chi2(8)    =   4.32
                                                        Prob > chi2   = 0.8272
Log likelihood = -238.07469                             Pseudo R2     = 0.0090

------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |   .0001538   .0085639     0.02   0.986    -.0166311    .0169387
  Client_Married |   .0495006     .21405     0.23   0.817    -.3700297     .469031
Client_Education |  -.0146394     .01661    -0.88   0.378    -.0471945    .0179157
```

```
         HH_Size |   -.047596    .037897    -1.26   0.209    -.1218727    .0266807
       HH_Income |   .0000133   .0000144     0.92   0.356    -.0000149    .0000415
          muslim |  -.0325982   .1467714    -0.22   0.824    -.3202648    .2550683
    Hindu_SC_Kat |  -.1099977   .2152225    -0.51   0.609     -.531826    .3118307
         Treated |   .1751011   .1418766     1.23   0.217    -.1029719    .4531741
           _cons |  -.8526965   .4591759    -1.86   0.063    -1.752665    .0472717
------------------------------------------------------------------------------

. outreg2 using p6_table, tex(frag) append addtext(Model, Probit)
p6_table.tex
dir : seeout

. logit taken_new Client_Age `covariates'

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.09541
Iteration 2:   log likelihood = -238.07838
Iteration 3:   log likelihood = -238.07838

Logistic regression                                     Number of obs =    532
                                                        LR chi2(8)    =   4.31
                                                        Prob > chi2   = 0.8280
Log likelihood = -238.07838                             Pseudo R2     = 0.0090

------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |   -.000382   .0156578    -0.02   0.981    -.0310707    .0303066
  Client_Married |   .0930805   .3876506     0.24   0.810    -.6667008    .8528617
Client_Education |  -.0275642   .0299653    -0.92   0.358    -.0862951    .0311666
         HH_Size |  -.0854032   .0693938    -1.23   0.218    -.2214126    .0506061
       HH_Income |   .0000223   .0000253     0.88   0.377    -.0000272    .0000718
          muslim |  -.0532995   .2633343    -0.20   0.840    -.5694252    .4628263
    Hindu_SC_Kat |  -.2078404     .39463    -0.53   0.598     -.981301    .5656203
         Treated |   .3186872   .2585782     1.23   0.218    -.1881167    .8254911
           _cons |  -1.374046   .8354666    -1.64   0.100     -3.01153    .2634387
------------------------------------------------------------------------------

. outreg2 using p6_table, tex(frag) append addtext(Model, Logit)
```

```
p6_table.tex
dir : seeout

.
.
. ****************************************************************************

. * problem #7 - mean partial derivatives of Client_Age

. ****************************************************************************

.
. * LPM

. regress taken_new Client_Age ‘covariates’

      Source |       SS           df       MS      Number of obs   =       532
-------------+----------------------------------   F(8, 523)       =      0.52
       Model |  .586701556         8  .073337695   Prob > F        =    0.8405
    Residual |  73.5242007       523  .140581646   R-squared       =    0.0079
-------------+----------------------------------   Adj R-squared   =   -0.0073
       Total |  74.1109023       531  .139568554   Root MSE        =    .37494

------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |  -.0000283   .0021572    -0.01   0.990     -.004266    .0042095
  Client_Married |   .0117141   .0529309     0.22   0.825    -.0922693    .1156975
Client_Education |  -.0036931   .0041228    -0.90   0.371    -.0117924    .0044062
         HH_Size |  -.0113264   .0093055    -1.22   0.224    -.0296072    .0069544
       HH_Income |   3.14e-06   3.68e-06     0.85   0.393    -4.08e-06    .0000104
          muslim |  -.0075629   .0366854    -0.21   0.837    -.0796318    .0645061
    Hindu_SC_Kat |  -.0274764   .0526227    -0.52   0.602    -.1308542    .0759015
         Treated |   .0426031   .0346907     1.23   0.220    -.0255471    .1107532
           _cons |   .1993902   .1142909     1.74   0.082    -.0251353    .4239158
------------------------------------------------------------------------------

. * mean partial derivative is just the LPM coefficient = -.0000283
```

```
.
. * probit

. * part a - using dprobit

. dprobit taken_new Client_Age `covariates'

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
Iteration 3:   log likelihood = -238.07469

Probit regression, reporting marginal effects           Number of obs =    532
                                                        LR chi2(8)    =   4.32
                                                        Prob > chi2   = 0.8272
Log likelihood = -238.07469                             Pseudo R2     = 0.0090

------------------------------------------------------------------------------
taken_~w |      dF/dx   Std. err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
Clien~ge |   .0000382   .0021261     0.02   0.986   34.8947  -.004129  .004205
Cl~rried*|   .0120563   .0511232     0.23   0.817   .892857  -.088143  .112256
Client~n |  -.0036344   .0041208    -0.88   0.378   6.35338  -.011711  .004442
 HH_Size |  -.0118163   .0093894    -1.26   0.209   5.30827  -.030219  .006587
HH_Inc~e |   3.30e-06   3.57e-06     0.92   0.356   6096.04  -3.7e-06   .00001
  muslim*|  -.0080417    .035974    -0.22   0.824   .300752   -.07855  .062466
Hi~C_Kat*|  -.0261864    .049033    -0.51   0.609   .116541  -.122289  .069917
 Treated*|   .0422445   .0331814     1.23   0.217   .665414   -.02279  .107279
---------+--------------------------------------------------------------------
  obs. P |   .1672932
 pred. P |   .1650313  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0

. * mean partial is the value that dprobit outputs = .0000382

.
. * part b - analytical derivative
```

```
. probit taken_new Client_Age ‘covariates’

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
Iteration 3:   log likelihood = -238.07469

Probit regression                                       Number of obs =    532
                                                        LR chi2(8)    =   4.32
                                                        Prob > chi2   = 0.8272
Log likelihood = -238.07469                             Pseudo R2     = 0.0090

------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |   .0001538   .0085639     0.02   0.986    -.0166311    .0169387
  Client_Married |   .0495006     .21405     0.23   0.817    -.3700297     .469031
Client_Education |  -.0146394     .01661    -0.88   0.378    -.0471945    .0179157
         HH_Size |   -.047596    .037897    -1.26   0.209    -.1218727    .0266807
       HH_Income |   .0000133   .0000144     0.92   0.356    -.0000149    .0000415
          muslim |  -.0325982   .1467714    -0.22   0.824    -.3202648    .2550683
    Hindu_SC_Kat |  -.1099977   .2152225    -0.51   0.609     -.531826    .3118307
         Treated |   .1751011   .1418766     1.23   0.217    -.1029719    .4531741
           _cons |  -.8526965   .4591759    -1.86   0.063    -1.752665    .0472717
------------------------------------------------------------------------------------

. * get the linear prediction based on probit

. predict taken_new_hat_xb, xb

. * using formula phi(xb)*b_j

. gen Client_Age_Partial_a = normalden(taken_new_hat_xb) * e(b)[1,1]

. summarize Client_Age_Partial_a

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
```

```
Client_Age~a |        532    .0000382    5.01e-06   .0000224   .0000546

. * mean partial derivative is the mean = .0000382

.
. * part c - numerically calculating marginal effects

. * predict probability based on probit

. predict taken_new_hat_probit, pr

. gen taken_new_hat_probit_epsilon = normal(taken_new_hat_xb + 0.001*e(b)[1,1])

.
. * compute numerical derivative

. gen Client_Age_Partial_n =  (taken_new_hat_probit_epsilon - taken_new_hat_probit) / 0.001

. summarize Client_Age_Partial_n

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
Client_Age~n |        532    .0000386    9.50e-06   .0000149   .0000596

. * mean partial derivative is the mean = 0

.
. * part d - using margins

. probit taken_new Client_Age `covariates'

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
Iteration 3:   log likelihood = -238.07469

Probit regression                                       Number of obs =    532
                                                        LR chi2(8)    =   4.32
                                                        Prob > chi2   = 0.8272
```

```
Log likelihood = -238.07469                              Pseudo R2     = 0.0090

------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |   .0001538   .0085639     0.02   0.986    -.0166311    .0169387
  Client_Married |   .0495006     .21405     0.23   0.817    -.3700297     .469031
Client_Education |  -.0146394     .01661    -0.88   0.378    -.0471945    .0179157
         HH_Size |   -.047596    .037897    -1.26   0.209    -.1218727    .0266807
       HH_Income |   .0000133   .0000144     0.92   0.356    -.0000149    .0000415
          muslim |  -.0325982   .1467714    -0.22   0.824    -.3202648    .2550683
    Hindu_SC_Kat |  -.1099977   .2152225    -0.51   0.609     -.531826    .3118307
         Treated |   .1751011   .1418766     1.23   0.217    -.1029719    .4531741
           _cons |  -.8526965   .4591759    -1.86   0.063    -1.752665    .0472717
------------------------------------------------------------------------------------

. margins , dydx(Client_Age) atmeans

Conditional marginal effects                                Number of obs = 532
Model VCE: OIM

Expression: Pr(taken_new), predict()
dy/dx wrt:  Client_Age
At: Client_Age       = 34.89474 (mean)
    Client_Married   = .8928571 (mean)
    Client_Education = 6.353383 (mean)
    HH_Size          = 5.308271 (mean)
    HH_Income        = 6096.039 (mean)
    muslim           = .3007519 (mean)
    Hindu_SC_Kat     = .1165414 (mean)
    Treated          = .6654135 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  Client_Age |   .0000382   .0021261     0.02   0.986    -.0041289    .0042052
------------------------------------------------------------------------------
```

```
. * mean partial derivative is the mean =  0.0000382

.
. * logit

. logit taken_new Client_Age `covariates'

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.09541
Iteration 2:   log likelihood = -238.07838
Iteration 3:   log likelihood = -238.07838

Logistic regression                                     Number of obs =    532
                                                        LR chi2(8)    =   4.31
                                                        Prob > chi2   = 0.8280
Log likelihood = -238.07838                             Pseudo R2     = 0.0090

---------------------------------------------------------------------------------
      taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
----------------+----------------------------------------------------------------
     Client_Age |   -.000382   .0156578    -0.02   0.981    -.0310707    .0303066
 Client_Married |   .0930805   .3876506     0.24   0.810    -.6667008    .8528617
Client_Education |  -.0275642   .0299653    -0.92   0.358    -.0862951    .0311666
        HH_Size |  -.0854032   .0693938    -1.23   0.218    -.2214126    .0506061
      HH_Income |   .0000223   .0000253     0.88   0.377    -.0000272    .0000718
         muslim |  -.0532995   .2633343    -0.20   0.840    -.5694252    .4628263
   Hindu_SC_Kat |  -.2078404     .39463    -0.53   0.598     -.981301    .5656203
        Treated |   .3186872   .2585782     1.23   0.218    -.1881167    .8254911
          _cons |  -1.374046   .8354666    -1.64   0.100     -3.01153    .2634387
---------------------------------------------------------------------------------

. margins , dydx(Client_Age) atmeans

Conditional marginal effects                               Number of obs = 532
Model VCE: OIM

Expression: Pr(taken_new), predict()
dy/dx wrt:  Client_Age
At: Client_Age       = 34.89474 (mean)
```

```
    Client_Married   = .8928571 (mean)
    Client_Education = 6.353383 (mean)
    HH_Size          = 5.308271 (mean)
    HH_Income        = 6096.039 (mean)
    muslim           = .3007519 (mean)
    Hindu_SC_Kat     = .1165414 (mean)
    Treated          = .6654135 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  Client_Age |  -.0000525   .0021521    -0.02   0.981    -.0042706    .0041656
------------------------------------------------------------------------------

. * mean partial derivative is the mean =   -.0000525

.
. ********************************************************************************

. * problem #8 - LPM with quadratic age

. ********************************************************************************

.
. * baseline

. regress taken_new `covariates', robust

Linear regression                               Number of obs     =        532
                                                F(7, 524)         =       0.64
                                                Prob > F          =     0.7206
                                                R-squared         =     0.0079
                                                Root MSE          =     .37458

----------------------------------------------------------------------------------
                 |               Robust
       taken_new | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------+----------------------------------------------------------------
```

```
  Client_Married |   .0117163   .0518345     0.23   0.821    -.0901127    .1135453
Client_Education |  -.0036812    .003937    -0.94   0.350    -.0114154     .004053
         HH_Size |  -.0113324   .0092732    -1.22   0.222    -.0295496    .0068847
       HH_Income |   3.14e-06   3.71e-06     0.85   0.398    -4.15e-06    .0000104
          muslim |  -.0075461   .0363378    -0.21   0.836    -.0789317    .0638396
    Hindu_SC_Kat |  -.0275317   .0509772    -0.54   0.589    -.1276765    .0726131
         Treated |   .0426362   .0335181     1.27   0.204      -.02321    .1084825
           _cons |   .1983494   .0779066     2.55   0.011     .0453018     .351397
---------------------------------------------------------------------------------

. outreg2 using p8_table, tex(frag) replace
p8_table.tex
dir : seeout

.
. * create quadratic transformations of age

. gen Client_Age_2 = Client_Age^2

. gen Client_Age_3 = Client_Age^3

. gen Client_Age_4 = Client_Age^4

.
. * estimate lpm with quadratic transformations of age

. regress taken_new Client_Age Client_Age_2 Client_Age_3 Client_Age_4 `covariates', robust

Linear regression                               Number of obs     =        532
                                                F(11, 520)        =      26.51
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0311
                                                Root MSE          =      .3716

---------------------------------------------------------------------------------
                |               Robust
      taken_new | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
----------------+----------------------------------------------------------------
     Client_Age |  -.4933651   .1676673    -2.94   0.003    -.8227536   -.1639766
```

```
    Client_Age_2 |   .0198727   .0062055     3.20   0.001     .0076817    .0320637
    Client_Age_3 |  -.0003363   .0000968    -3.47   0.001    -.0005265   -.0001461
    Client_Age_4 |   2.01e-06   5.35e-07     3.76   0.000     9.60e-07    3.06e-06
  Client_Married |   .0151682   .0552224     0.27   0.784    -.0933182    .1236546
Client_Education |  -.0031541   .0041087    -0.77   0.443    -.0112258    .0049176
         HH_Size |  -.0088907   .0092416    -0.96   0.336    -.0270461    .0092647
       HH_Income |   3.65e-06   3.70e-06     0.99   0.325    -3.62e-06    .0000109
           muslim |  -.0128246   .0361322    -0.35   0.723    -.0838076    .0581584
    Hindu_SC_Kat |  -.0333433   .0506817    -0.66   0.511    -.1329092    .0662227
         Treated |   .0464548   .0334116     1.39   0.165    -.0191835    .1120931
           _cons |   4.517748   1.604795     2.82   0.005     1.365071    7.670425
----------------------------------------------------------------------------------

. outreg2 using p8_table, tex(frag) append
p8_table.tex
dir : seeout

. predict taken_new_hat_lpm_q
(option xb assumed; fitted values)

.
. * check for observations outside 0, 1

. count if taken_new_hat_lpm_q > 1
  1

. count if taken_new_hat_lpm_q < 0
  5

. histogram taken_new_hat_lpm_q
(bin=23, start=-.16169128, width=.05206823)

. graph export p8_figure.png, replace
file /Users/alexandervonhafften/Documents/UW Madison/problem_sets/econ_717a/ps1/p8_figure.png saved as PNG format

.
. * create quadratic transformations with epsilon

. gen Client_Age_epsilon_2 = Client_Age_epsilon^2
```

```
Client_Age_epsilon not found
r(111);

. gen Client_Age_epsilon_3 = Client_Age_epsilon^3
Client_Age_epsilon not found
r(111);

. gen Client_Age_epsilon_4 = Client_Age_epsilon^4
Client_Age_epsilon not found
r(111);

. regress taken_new Client_Age_epsilon Client_Age_epsilon_2 Client_Age_epsilon_3 Client_Age_epsilon_4 `covariates'
variable Client_Age_epsilon not found
r(111);

. predict taken_new_hat_lpm_q_epsilon
(option xb assumed; fitted values)

.
. * numerically compute derivative

. gen Client_Age_Partial_q_n =  (taken_new_hat_lpm_q - taken_new_hat_lpm_q_epsilon) / 0.001

. summarize Client_Age_Partial_q_n

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
Client_A~q_n |        532           0           0          0          0

. * mean is 2.32e-07

.
. ********************************************************************************

. * problem #9 - LRI

. ********************************************************************************

.
```

```
. * baseline probit

. probit taken_new Client_Age ‘covariates’

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
Iteration 3:   log likelihood = -238.07469

Probit regression                                       Number of obs =    532
                                                        LR chi2(8)    =   4.32
                                                        Prob > chi2   = 0.8272
Log likelihood = -238.07469                             Pseudo R2     = 0.0090

------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |   .0001538   .0085639     0.02   0.986    -.0166311    .0169387
  Client_Married |   .0495006     .21405     0.23   0.817    -.3700297     .469031
Client_Education |  -.0146394     .01661    -0.88   0.378    -.0471945    .0179157
         HH_Size |   -.047596    .037897    -1.26   0.209    -.1218727    .0266807
       HH_Income |   .0000133   .0000144     0.92   0.356    -.0000149    .0000415
          muslim |  -.0325982   .1467714    -0.22   0.824    -.3202648    .2550683
    Hindu_SC_Kat |  -.1099977   .2152225    -0.51   0.609     -.531826    .3118307
         Treated |   .1751011   .1418766     1.23   0.217    -.1029719    .4531741
           _cons |  -.8526965   .4591759    -1.86   0.063    -1.752665    .0472717
------------------------------------------------------------------------------------

. * ll_hat is  -238.07469

.
. * probit only with constant

. probit taken_new

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -240.23429

Probit regression                                       Number of obs =    532
```

```
                                                        LR chi2(0)    =   0.00
                                                        Prob > chi2   =      .
Log likelihood = -240.23429                             Pseudo R2     = 0.0000

------------------------------------------------------------------------------
   taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |  -.9649168   .0646097   -14.93   0.000    -1.091549   -.8382842
------------------------------------------------------------------------------

. * ll_0 is -240.23429

.
. * lri = 1 - ll_hat/ll_0 = 0.00898955765

.
. ********************************************************************************

. * problem #10 - Prediction rate

. ********************************************************************************

.
. * using cutoff of 50 percent

. gen predicted_over_50 =  taken_new_hat_probit > .5

. tab taken_new predicted_over_50, nofreq cell

 Has taken |
a new loan |
    in the |
    last 4 | predicted_
    months |  over_50
 (midline) |         0 |     Total
-----------+-----------+----------
         0 |     83.27 |     83.27
         1 |     16.73 |     16.73
-----------+-----------+----------
```

```
     Total |     100.00 |     100.00

.
. * using unconditional probability as cutoff

. tab taken_new

Has taken a |
new loan in |
 the last 4 |
     months |
  (midline) |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        443       83.27       83.27
          1 |         89       16.73      100.00
------------+-----------------------------------
      Total |        532      100.00

. * unconditional probability = .1673

. gen predicted_over_up =  taken_new_hat_probit > .1673

. tab taken_new predicted_over_up, nofreq cell

 Has taken |
a new loan |
    in the |
    last 4 |
    months |   predicted_over_up
 (midline) |         0          1 |     Total
-----------+----------------------+----------
         0 |     42.67      40.60 |     83.27
         1 |      6.02      10.71 |     16.73
-----------+----------------------+----------
     Total |     48.68      51.32 |    100.00

.
. ********************************************************************************
```

```
. * problem #11 - In sample vs. out-of-sample prediction

. ****************************************************************************

.
. gen estimation_sample = imidlineid < 1400

.
. * estimate probit on subsample

. probit taken_new Client_Age ‘covariates’ if estimation_sample

Iteration 0:   log likelihood = -117.67911
Iteration 1:   log likelihood = -115.82624
Iteration 2:   log likelihood = -115.80635
Iteration 3:   log likelihood = -115.80635

Probit regression                                       Number of obs =    266
                                                        LR chi2(8)    =   3.75
                                                        Prob > chi2   = 0.8793
Log likelihood = -115.80635                             Pseudo R2     = 0.0159

------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |   .0095479   .0132631     0.72   0.472    -.0164472    .0355431
  Client_Married |   .1217882    .317054     0.38   0.701    -.4996262    .7432027
Client_Education |   .0020014   .0246685     0.08   0.935    -.0463478    .0503507
         HH_Size |   .0450808   .0526267     0.86   0.392    -.0580656    .1482272
       HH_Income |   4.91e-06     .00002     0.25   0.806    -.0000342    .0000441
          muslim |   .0064283   .2123982     0.03   0.976    -.4098646    .4227212
    Hindu_SC_Kat |  -.5059869    .387024    -1.31   0.191     -1.26454    .2525661
         Treated |   .0434882   .2025915     0.21   0.830     -.353584    .4405603
           _cons |  -1.708002   .7275063    -2.35   0.019    -3.133888    -.282116
------------------------------------------------------------------------------------

. predict taken_new_hat_probit_11, pr

.
```

```
. * using cutoff of 50 percent

. gen predicted_over_50_11 =  taken_new_hat_probit_11 > .5

. tab taken_new predicted_over_50_11 if estimation_sample, cell nofreq

 Has taken |
a new loan |
    in the |
    last 4 | predicted_
    months | over_50_11
 (midline) |         0 |     Total
-----------+-----------+----------
         0 |     83.83 |     83.83
         1 |     16.17 |     16.17
-----------+-----------+----------
     Total |    100.00 |    100.00

. tab taken_new predicted_over_50_11 if !estimation_sample, cell  nofreq

 Has taken |
a new loan |
    in the |
    last 4 | predicted_
    months | over_50_11
 (midline) |         0 |     Total
-----------+-----------+----------
         0 |     82.71 |     82.71
         1 |     17.29 |     17.29
-----------+-----------+----------
     Total |    100.00 |    100.00

.
. * using unconditional probability as cutoff

. gen predicted_over_up_11 =  taken_new_hat_probit_11 > .1673

. tab taken_new predicted_over_up_11 if estimation_sample, cell nofreq
```

```
 Has taken |
a new loan |
    in the |
    last 4 |
    months | predicted_over_up_11
 (midline) |         0          1 |     Total
-----------+----------------------+----------
         0 |     48.12      35.71 |     83.83
         1 |      8.27       7.89 |     16.17
-----------+----------------------+----------
     Total |     56.39      43.61 |    100.00

. tab taken_new predicted_over_up_11 if !estimation_sample, cell nofreq

 Has taken |
a new loan |
    in the |
    last 4 |
    months | predicted_over_up_11
 (midline) |         0          1 |     Total
-----------+----------------------+----------
         0 |     44.36      38.35 |     82.71
         1 |     11.28       6.02 |     17.29
-----------+----------------------+----------
     Total |     55.64      44.36 |    100.00

.
. ********************************************************************************

. * problem #12 - Interaction terms

. ********************************************************************************

.
. probit taken_new Client_Age ‘covariates’

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
```

```
Iteration 3:   log likelihood = -238.07469

Probit regression                                       Number of obs =    532
                                                        LR chi2(8)    =   4.32
                                                        Prob > chi2   = 0.8272
Log likelihood = -238.07469                             Pseudo R2     = 0.0090

------------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
      Client_Age |   .0001538   .0085639     0.02   0.986    -.0166311    .0169387
  Client_Married |   .0495006     .21405     0.23   0.817    -.3700297     .469031
Client_Education |  -.0146394     .01661    -0.88   0.378    -.0471945    .0179157
         HH_Size |   -.047596    .037897    -1.26   0.209    -.1218727    .0266807
       HH_Income |   .0000133   .0000144     0.92   0.356    -.0000149    .0000415
          muslim |  -.0325982   .1467714    -0.22   0.824    -.3202648    .2550683
    Hindu_SC_Kat |  -.1099977   .2152225    -0.51   0.609     -.531826    .3118307
         Treated |   .1751011   .1418766     1.23   0.217    -.1029719    .4531741
           _cons |  -.8526965   .4591759    -1.86   0.063    -1.752665    .0472717
------------------------------------------------------------------------------------

. outreg2 using p12_table, tex(frag) replace
p12_table.tex
dir : seeout

.
. gen married_muslim = Client_Married * muslim

. probit taken_new Client_Age `covariates' married_muslim

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -237.89689
Iteration 2:   log likelihood = -237.89262
Iteration 3:   log likelihood = -237.89262

Probit regression                                       Number of obs =    532
                                                        LR chi2(9)    =   4.68
                                                        Prob > chi2   = 0.8610
Log likelihood = -237.89262                             Pseudo R2     = 0.0097
```

```
------------------------------------------------------------------------------
   taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  Client_Age |   .0009312   .0086796     0.11   0.915    -.0160805    .0179429
Client_Married |   .1556674   .2806756     0.55   0.579    -.3944468    .7057815
Client_Education |  -.0153534   .0166616    -0.92   0.357    -.0480094    .0173027
     HH_Size |  -.0502379   .0381311    -1.32   0.188    -.1249735    .0244977
   HH_Income |   .0000126   .0000145     0.87   0.383    -.0000157     .000041
      muslim |   .2059269    .420573     0.49   0.624    -.6183811    1.030235
 Hindu_SC_Kat |  -.1142815   .2152731    -0.53   0.596     -.536209     .307646
     Treated |   .1819567   .1423657     1.28   0.201    -.0970751    .4609884
married_muslim |  -.2709185   .4481227    -0.60   0.545    -1.149223    .6073858
       _cons |  -.9589165   .4938953    -1.94   0.052    -1.926934    .0091005
------------------------------------------------------------------------------

. outreg2 using p12_table, tex(frag) append
p12_table.tex
dir : seeout

.
. ****************************************************************************

. * problem #13 - Interaction effects

. ****************************************************************************

.
. * compute interaction effect without accounting for terms in Ai and Norton (2003)

. margins , dydx(married_muslim)

Average marginal effects                                   Number of obs = 532
Model VCE: OIM

Expression: Pr(taken_new), predict()
dy/dx wrt:  married_muslim

------------------------------------------------------------------------------
```

```
              |            Delta-method
              |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
--------------+----------------------------------------------------------------
married_muslim |  -.0671875   .1110788    -0.60   0.545     -.284898    .1505229
-------------------------------------------------------------------------------

. * interaction effect estimate is -.0671875

.
. * compute interaction effect by hand accounting for terms in Ai and Norton (2003)

. * follows logic from lecture notes

. predict index_hat, xb

. * predicted index with both dummies zero

. gen index_hat_0 = index_hat - Client_Married * e(b)[1,2] - muslim * e(b)[1,6] - married_muslim *e(b)[1, 9]

. * predicted index with both married one and muslim zero

. gen index_hat_01 = index_hat_0 + e(b)[1,2]

. * predicted index with both muslim one and married zero

. gen index_hat_02 = index_hat_0 + e(b)[1,6]

. * predicted index with both dummies zero

. gen index_hat_012 = index_hat_0 + e(b)[1,2] + e(b)[1,6] + e(b)[1,9]

. gen finite_difference = (normal(index_hat_012) - normal(index_hat_02)) - (normal(index_hat_01) - normal(index_hat_0))

. summarize finite_difference

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
finite_dif~e |        532   -.0656881    .0090154  -.0940148  -.0364692
```

```
. * interaction effect estimate is -.0656881

.
. ****************************************************************************

. * problem #14 - Interaction effects variance

. ****************************************************************************

.
. * see summarize table from problem #13.

.
. ****************************************************************************

. * problem #15 - Heteroskedasticity test

. ****************************************************************************

.
. * compute residuals

. regress taken_new Client_Age `covariates'

      Source |       SS           df       MS      Number of obs   =       532
-------------+----------------------------------   F(8, 523)       =      0.52
       Model |  .586701556         8  .073337695   Prob > F        =    0.8405
    Residual |  73.5242007       523  .140581646   R-squared       =    0.0079
-------------+----------------------------------   Adj R-squared   =   -0.0073
       Total |  74.1109023       531  .139568554   Root MSE        =    .37494

------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |  -.0000283   .0021572    -0.01   0.990     -.004266    .0042095
  Client_Married |   .0117141   .0529309     0.22   0.825    -.0922693    .1156975
Client_Education |  -.0036931   .0041228    -0.90   0.371    -.0117924    .0044062
         HH_Size |  -.0113264   .0093055    -1.22   0.224    -.0296072    .0069544
       HH_Income |   3.14e-06   3.68e-06     0.85   0.393    -4.08e-06    .0000104
```

```
         muslim |  -.0075629   .0366854    -0.21   0.837    -.0796318    .0645061
    Hindu_SC_Kat |  -.0274764   .0526227    -0.52   0.602    -.1308542    .0759015
         Treated |   .0426031   .0346907     1.23   0.220    -.0255471    .1107532
           _cons |   .1993902   .1142909     1.74   0.082    -.0251353    .4239158
------------------------------------------------------------------------------

. predict residuals_p, residuals

.
. * regress squared residuals on usual covariates.

. gen residuals_p_2 = residuals_p^2

. regress residuals_p_2 Client_Age ‘covariates’

      Source |       SS           df       MS      Number of obs   =       532
-------------+----------------------------------   F(8, 523)       =      0.56
       Model |  .271068132         8  .033883517   Prob > F        =    0.8130
    Residual |  31.8045385       523  .060811737   R-squared       =    0.0085
-------------+----------------------------------   Adj R-squared   =   -0.0067
       Total |  32.0756066       531  .060406039   Root MSE        =     .2466

------------------------------------------------------------------------------
   residuals_p_2 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------
      Client_Age |   .0001681   .0014188     0.12   0.906     -.002619    .0029553
  Client_Married |   .0077442   .0348128     0.22   0.824    -.0606459    .0761344
Client_Education |   -.002132   .0027116    -0.79   0.432    -.0074589    .0031949
         HH_Size |   -.007861   .0061203    -1.28   0.200    -.0198843    .0041623
       HH_Income |   2.51e-06   2.42e-06     1.04   0.299    -2.24e-06    7.27e-06
          muslim |  -.0063261   .0241281    -0.26   0.793     -.053726    .0410738
    Hindu_SC_Kat |  -.0167612   .0346101    -0.48   0.628    -.0847531    .0512306
         Treated |   .0281959   .0228161     1.24   0.217    -.0166267    .0730184
           _cons |   .1504638   .0751694     2.00   0.046     .0027929    .2981348
------------------------------------------------------------------------------

. outreg2 using p15_table, tex(frag) replace
p15_table.tex
dir : seeout
```

```
.
.
. ********************************************************************************

. * problem #16 - hetprob

. ********************************************************************************

.
. hetprob taken_new Client_Age `covariates', het(Client_Age Client_Education)

Fitting probit model:

Iteration 0:   log likelihood = -240.23429
Iteration 1:   log likelihood = -238.07858
Iteration 2:   log likelihood = -238.07469
Iteration 3:   log likelihood = -238.07469

Fitting full model:

Iteration 0:   log likelihood = -238.07469  (not concave)
Iteration 1:   log likelihood = -237.99107
Iteration 2:   log likelihood = -237.64163
Iteration 3:   log likelihood = -237.57256
Iteration 4:   log likelihood = -237.11403
Iteration 5:   log likelihood =  -236.9148
Iteration 6:   log likelihood = -236.69574
Iteration 7:   log likelihood = -236.60898
Iteration 8:   log likelihood = -236.59195
Iteration 9:   log likelihood = -236.59073
Iteration 10:  log likelihood =  -236.5907
Iteration 11:  log likelihood =  -236.5907

Heteroskedastic probit model                    Number of obs     =        532
                                                Zero outcomes     =        443
                                                Nonzero outcomes  =         89

                                                Wald chi2(8)      =       2.19
```

```
Log likelihood = -236.5907                      Prob > chi2       =     0.9745

--------------------------------------------------------------------------------
       taken_new | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+--------------------------------------------------------------
taken_new        |
      Client_Age |  -.1123273   .1374995    -0.82   0.414    -.3818214    .1571668
  Client_Married |   .1288146   .8458704     0.15   0.879    -1.529061     1.78669
Client_Education |  -.3106294   .2555354    -1.22   0.224    -.8114697    .1902108
         HH_Size |  -.2261131   .1970247    -1.15   0.251    -.6122744    .1600482
       HH_Income |   .0000693   .0000723     0.96   0.338    -.0000723    .0002109
          muslim |  -.1792461   .5843143    -0.31   0.759    -1.324481    .9659888
    Hindu_SC_Kat |  -.3442778   .9618559    -0.36   0.720    -2.229481    1.540925
         Treated |   .9149664   .7866622     1.16   0.245    -.6268631    2.456796
           _cons |    1.71513    3.57286     0.48   0.631    -5.287547    8.717807
-----------------+--------------------------------------------------------------
lnsigma          |
      Client_Age |   .0285154   .0196403     1.45   0.147    -.0099788    .0670096
Client_Education |   .0694184    .047379     1.47   0.143    -.0234426    .1622795
--------------------------------------------------------------------------------
LR test of lnsigma=0: chi2(2) = 2.97                      Prob > chi2 = 0.2267

. outreg2 using p16_table, tex(frag) replace
p16_table.tex
dir : seeout

.
. log close
      name:  <unnamed>
       log:  /Users/alexandervonhafften/Documents/UW Madison/problem_sets/econ_717a/ps1/analysis.smcl
  log type:  smcl
 closed on:  12 Feb 2022, 13:29:57
-------------------------------------------------------------------------------------------------------------------------------------
```