GVPT728 HW#4

Yael Beshaw

2025-01-10

Download Data of Interest

```
library(tidyverse)
library(modelsummary)
socpoc <- dataverse::get_dataframe_by_name(
  filename = "socpocAPSR.csv",
  .f=read_csv,
  dataset = "10.7910/DVN/CRPAA8&version=1.0",
  server = "dataverse.harvard.edu")|>
  select(-1)
```

Question 1

Estimate the following bivariate model using a linear probability model and a logit or probit model

```
lpm <- lm(supply_dummy ~ own , data = socpoc)

logit <- glm(supply_dummy ~ own , data = socpoc,
             family = binomial)

probit <- glm(supply_dummy ~ own , data = socpoc,
             family = binomial(link= "probit"))
```

Comparision of Models

```
compare<-list("Linear Probability" = lpm, "Logit" = logit, "Probit"= probit)
modelsummary(compare, output = "huxtable")
```

Question 1a

What is the estimated effect of home-ownership on the probability that a respondent supports reduced restrictions on building new housing in this model? (state your answer in terms of the difference in % probability). Note any* differences between your three models.*

Linear Probability Model

The estimated effect of home-ownership on the probability that a respondent supports reduced restrictions on building new housing in the Linear Probability Model is 31.2% lower than that of non-homeownership.

	Linear Probability	Logit	Probit
(Intercept)	0.588	0.354	0.221
	(0.019)	(0.083)	(0.052)
own	-0.312	-1.321	-0.818
	(0.023)	(0.104)	(0.064)
Num.Obs.	1909	1909	1909
R2	0.090		
R2 Adj.	0.089		
AIC	2471.5	2358.3	2358.3
BIC	2488.2	2369.5	2369.5
Log.Lik.	-1232.772	-1177.172	-1177.172
F	187.723	162.830	165.621
RMSE	0.46	0.46	0.46

Logit Model

The estimated effect of home-ownership that a respondent supports reduced restrictions on building new housing in the Logit Model 31.2% less compared to non-homeownership.

```
coefs<-coef(logit)
logit_pred0 <- coefs["(Intercept)"]
logit_pred1 <- coefs["(Intercept)"] + coefs['own'] * 1

pred_prob0 <- 1/(1+exp(-logit_pred0))
pred_prob1 <-1/(1+exp(-logit_pred1))

pred_prob0-pred_prob1
```

```
## (Intercept)
##   0.3120736
```

Probit Model

Similar to the other two models, the estimated effect of home-ownership that a respondent supports reduced restrictions on building new housing in the Probit Model 31.2% less compared to non-homeownership.

```
coefs<-coef(probit)

probit_pred0 <- coefs["(Intercept)"]
probit_pred1 <- coefs["(Intercept)"] + coefs['own'] * 1

probit_prob0 <- probit$family$linkinv(probit_pred0)
probit_prob1 <- probit$family$linkinv(probit_pred1)
```

```r
probit_prob0 - probit_prob1
```

```
## (Intercept)
##   0.3120736
```

Differences in the Models

The main differences in the models is that in our huxtable we see that that the linear probability model is the only one where the information for the intercept and the coefficents is directly interpretable. However, conducting predicited probabilities for both the logit and probit model, we see that we observe the same effect of predicted probability for a homeowner supporting reduced restrictions on building new housing. The logit and probit models both have the same BIC and AIC which is lower than that of the linear probabilities model. This indicates that the logit/probit may be a better model, however, the F statisitc for the probit is slightly higher than that of the logit, indicating that it may be the better fit overall.

Question 2

Estimate a linear probability and logit/probit model with the same dependent variable, but include additional controls for ideology, white (non-hispanic) and income:

```r
lpm_new <- lm(supply_dummy ~ own + ideology + whitenh + income , data = socpoc)

logit_new <- glm(supply_dummy ~ own + ideology + whitenh + income ,
                 data = socpoc,
                 family= binomial)

probit_new <- glm(supply_dummy ~ own + ideology + whitenh + income ,
                  data = socpoc,
                  family= binomial(link="probit"))
```

Comparision of Models

```r
compare<-list("Linear Probability" = lpm_new, "Logit" = logit_new,
              "Probit"= probit_new)

modelsummary(compare, output = "huxtable")
```

Question 2a

What is the marginal effect of homeownership for a non-white “Extremely Conservative” respondent whose household income is less than $5,000 per year? Compare and contrast the estimates from the linear probability model and the logit/probit model.*

Linear Probability Model

The marginal effect of homeownership for a non-white “Extremely Conservative” respondent whose household income is less than $5,000 per year on on the probability that the respondent supports reduced restrictions on building new housing in the Linear Probability Model is 27.4% lower than that if they were a non-homeowner.

```r
#marginal effect of linear probability model: -0.274
coef(lpm_new)["own"]
```

```
##        own
## -0.2744484
```

	Linear Probability	Logit	Probit
(Intercept)	0.517	0.017	0.015
	(0.039)	(0.182)	(0.111)
own	-0.274	-1.169	-0.723
	(0.025)	(0.114)	(0.070)
ideology	0.028	0.135	0.082
	(0.007)	(0.033)	(0.020)
whitenh	-0.085	-0.392	-0.238
	(0.023)	(0.106)	(0.065)
income	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Num.Obs.	1878	1878	1878
R2	0.106		
R2 Adj.	0.104		
AIC	2397.8	2287.3	2287.0
BIC	2431.0	2314.9	2314.7
Log.Lik.	-1192.887	-1138.627	-1138.496
RMSE	0.46	0.46	0.46

Logit

The marginal effect of homeownership for a non-white "Extremely Conservative" respondent whose household income is less than $5,000 per year on on the probability that the respondent supports reduced restrictions on building new housing in the Logit Model is 27.2% lower than that if they were a non-homeowner.

```
#marginal effect of logit model: -0.272
logit_effect_owner <-predict(logit_new,
                        newdata=data.frame(
                          own = 1,
                          income = 1,
                          whitenh = 0,
                          ideology = 1), type='response')

logit_effect_nonowner <-predict(logit_new,
                        newdata=data.frame(
                          own = 0,
                          income = 1,
                          whitenh = 0,
                          ideology = 1), type='response')
```

```
logit_effect_owner- logit_effect_nonowner
```

```
##          1
## -0.2723415
```

Probit

The marginal effect of homeownership for a non-white "Extremely Conservative" respondent whose household income is less than $5,000 per year on on the probability that the respondent supports reduced restrictions on building new housing in the Probit Model is 27.3% lower than that if they were a non-homeowner.

```
#marginal effect of probit model: -0.273
probit_effect_owner <-predict(probit_new,
                         newdata=data.frame(
                           own = 1,
                           income = 1,
                           whitenh = 0,
                           ideology = 1), type='response')

probit_effect_nonowner <-predict(probit_new,
                         newdata=data.frame(
                           own = 0,
                           income = 1,
                           whitenh = 0,
                           ideology = 1), type='response')

probit_effect_owner- probit_effect_nonowner
```

```
##        1
## -0.27284
```

Differences in the Models

The estimates are all around -27% across the three models. All three models show the same positive vs negative effect of each predictor. However, as seen in Question #1, the Probit model still has the lowest AIC and BIC out of the three with the highest logged likelihood, indicating that it is the model with the best fit.

Question 2b

Finally, estimate the average effect of homeownership across all observations using the observed values/counterfactual approach. Note any differences in your results.

Linear Probability Model

The average effect of homeownership is a is a 27.4% decrease in the probability of support of reduced restrictions on building new housing.

```
homeownership <- replace(lpm_new$model, "own", values=1)
non_homeownership <- replace(lpm_new$model, "own", values=0)

diffs<-predict(lpm_new, newdata=non_homeownership, type='response') -
  predict(lpm_new, newdata=homeownership, type='response')

mean(diffs)
```

```
## [1] 0.2744484
```

Logit

The average effect of homeownership is a is a 27.1% decrease in the probability of support of reduced restrictions on building new housing. This is slightly lower than the marginal effect of 27.2%

```
homeownership <- replace(logit_new$model, "own", values=1)
non_homeownership <- replace(logit_new$model, "own", values=0)

diffs<-predict(logit_new, newdata=non_homeownership, type='response') -
  predict(logit_new, newdata=homeownership, type='response')

mean(diffs)
```

```
## [1] 0.271065
```

Probit

The average effect of homeownership is a is a 27.2% decrease in the probability of support of reduced restrictions on building new housing. This is slightly lower than the marginal effect of 27.3%

```
homeownership <- replace(probit_new$model, "own", values=1)
non_homeownership <- replace(probit_new$model, "own", values=0)

diffs<-predict(probit_new, newdata=non_homeownership, type='response') -
  predict(probit_new, newdata=homeownership, type='response')

mean(diffs)
```

```
## [1] 0.2717926
```