Binary Choice Model

Kaden Grace Lucas Gurgel John Paul Lynn

PS 2 Presentation, March 7, 2023

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Summary of Data

Table: Summary Statistics

	Mean	SD	Min	Max
Good Health	0.70	0.46	0	1.00
Married	0.73	0.44	0	1.00
Hispanic	0.07	0.26	0	1.00
HH Income	45.26	64.34	0	1312.12
Retired	0.62	0.48	0	1.00
Yrs Education	11.90	3.30	0	17.00
Retired Spouse	0.39	0.49	0	1.00
Female	0.48	0.50	0	1.00
White	0.82	0.38	0	1.00
Observations	3206			

Logit and Probit Estimation

Table: Logit and Probit Estimation, Y = Bought Insurance

	Logit		Probit	
Bought Insurance				
Age	-0.007	(0.011)	-0.004	(0.007)
Good Health	0.318***	(0.092)	0.201***	(0.055)
HH Income	0.002***	(0.001)	0.001***	(0.000)
Yrs Education	0.118***	(0.014)	0.073***	(0.008)
Married	0.596***	(0.093)	0.373***	(0.056)
Hispanic	-0.820***	(0.196)	-0.478***	(0.110)
Constant	-2.141***	(0.724)	-1.326***	(0.444)
Observations	3206		3206	

Logit Estimation Interpretation

Logit Interpretation

- The logit coefficients need to be exponentiated before we can interpret them.
- Using our β, calculate the following: $e^{\beta} = i$, where the distance i is from 1 represents the change in the odds of receiving treatment.
- For example, if a person is recorded as having good health, our coefficient of 0.318 suggests that $e^{.318} = 1.3744$. This is .37 above 1, so having good health is interpreted as having a 37 percent increase in odds for buying insurance.
- All logit coefficients are interpreted this way. The Hispanic variable has an exponentiated value less than 1, so it represents a decrease in the odds of buying insurance.

Probit Estimation Interpretation

- Need to be careful interpreting probit results.
- They are not similar to the logit interpretations.
- Coefficients represent a change in the z-score of the standard normal distribution.
- For example, reporting good health will increase the z-score by 0.201.

Logit and Probit Estimation with `retire`

Table: Logit and Probit Estimation with `retire`, $Y =$ Bought Insurance

	Logit		Probit	
Bought Insurance				
Retired	0.197**	(0.084)	0.118**	(0.051)
Age	-0.015	(0.011)	-0.009	(0.007)
Good Health	0.312***	(0.092)	0.198***	(0.055)
HH Income	0.002***	(0.001)	0.001***	(0.000)
Yrs Education	0.114***	(0.014)	0.071***	(0.008)
Married	0.579***	(0.093)	0.362***	(0.056)
Hispanic	-0.810***	(0.196)	-0.473***	(0.110)
Constant	-1.716**	(0.749)	-1.069**	(0.458)
N	3206.000		3206.000	
ll	-1994.878		-1993.624	

Wald Test

Using the regression

$$
\begin{aligned}
ins_i = &\beta_0 + \beta_1 retire_i \\
&+ \beta_2 age_i \\
&+ \beta_3 hstatusg_i \\
&+ \beta_4 hhincome_i \\
&+ \beta_5 educyear_i \\
&+ \beta_6 married_i \\
&+ \beta_7 hisp_i \\
&+ \beta_8 age_i^2 \\
&+ \beta_9 age * female_i \\
&+ \beta_{10} age * chronic_i \\
&+ \beta_{11} age * white_i + \epsilon_i
\end{aligned}
$$

Test the null hypothesis: $H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$

Wald Test cont.

Test the null hypothesis: $H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$

- Yields χ^2_4 test statistic 7.45 and p-value 0.1141.
- This test statistic and p-value allows us to fail to reject the null hypothesis, since our p-value is above .1 (Not within in the 90 percent interval.
- Thus, we can not say with reasonable certainty that the interactions are significant.

Likelihood-Ratio Test

Using LR, test the same null hypothesis:
$H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$

- Yields test statistic 7.57 and p-value 0.1088.
- Results are similar to the Wald test, which is not surprising due to our large sample.
- Using the F-test requires calculating residuals, which we do not do in the logit model, so the F-test cannot be performed.

Predictive Classification

Table: Predictive Classification of Model

	$y = 1$	$y = 0$
$\hat{y} = 1$	**345**	308
$\hat{y} = 0$	896	**1657**

This model correctly classified 2002 of 3206 observations, or 62.45%.

Marginal Effects

Table: Various Marginal Effects

	MER	MEM	AME
Retired=1	0.035**	0.046**	0.043**
	(0.015)	(0.019)	(0.018)
Age	-0.003	-0.003	-0.003
	(0.002)	(0.003)	(0.002)
Good Health=1	0.054***	0.072***	0.068***
	(0.016)	(0.021)	(0.020)
HH Income	0.000***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Yrs Education	0.022***	0.027***	0.025***
	(0.004)	(0.003)	(0.003)
Married=1	0.094***	0.130***	0.124***
	(0.017)	(0.020)	(0.019)
Hispanic=1	-0.179***	-0.168***	-0.161***
	(0.038)	(0.034)	(0.034)

Interpretation of MER, MEM, and AME

- There are a few things to notice when comparing the three different logit tests when we examine them at the margin.
- The MER shows the lowest smallest effect for all coefficients.
- The MEM shows the highest effect for all coefficients.
- The AME coefficents are in between the effects of the MER and MEM.

Credit Market Discrimination

Initial Setup

- If we want study discrimination in the credit market, one strategy possible is to estimate the Average Partial Effect (APE) considering one characteristic X_{ij}, that is our point of interest.
- Suppose that our X_{ij} could be some "minority" characteristic, or race, or "bad credit history", and so on.
  - $X_{ij} = 1$, if "i" individual is a minority.
  - $X_{ij} = 0$, otherwise.

Credit Market Discrimination cont.

Initial Setup

- We still must have other K characteristics for each individual such as education, family background, wage, that are not the point of interest but they will serve to predict our probability function.
- Once we have the data set with a X vector containing all individuals characteristics, including the characteristics j that we want to study, and the outcome of interest y_i, where:
  - $y_i = 1$, if individual i's mortgage was denied or
  - $y_i = 0$, otherwise.

Credit Market Discrimination cont.

APE Regression

We can run our estimation in four steps:

- Step 1: Predicting the treatment group.

  $$Prob(y_i = 1/X_i, X_{ij} = 1) = F(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + ... + \hat{\beta}_j .1 + ... + \hat{\beta}_k X_{ik})$$

  for i=1,2,...,n.

- Step 2: Predicting the control group.

  $$Prob(y_i = 1/X_i, X_{ij} = 0) = F(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + ... + \hat{\beta}_j .0 + ... + \hat{\beta}_k X_{ik})$$

  for i=1,2,...,n.

Credit Market Discrimination cont.

- Step 3: Compute the $A\hat{P}E$ (average partial effects).

$$A\hat{P}E = \frac{1}{n}\sum_{i=1}^{n}[F(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + ... + \hat{\beta}_j .1 + ... + \hat{\beta}_k X_{ik})$$

$$-F(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + ... + \hat{\beta}_j .0 + ... + \hat{\beta}_k X_{ik})]$$

- Step 4: Test your $A\hat{P}E$.
  If $|A\hat{P}E| > 0$, it means that there is discrimination in the credit market.

Credit Market Discrimination cont.

Therefore, we want to test $H_0 : APE = 0$, to do so, we may use a t-test:

$$t = \frac{A\hat{P}E}{SE(A\hat{P}E)}$$

But, pay attention here, to calculate standard deviation of $A\hat{P}E$, we need to use delta method.
After follow this setup and these 4 steps, if you find that you do not have strong evidence to reject your null hypothesis, that means you may not infer that there is discrimination in the market for credit. However, if you reject the null hypothesis ($APE > 0$), thus we may say that there is some sort of discrimination in that market.