POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

Outline

1. Logit vs Scobit
2. Plotting Logits and Scobits
3. Example: *The Effect of Registration Laws and Education on U.S. Voter Turnout*
4. Example, continued: *Scobit: An Alternative Estimator to Logit*

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

0. Two approaches to binary response

A. Latent variables approach
- Assumes there exists a continuous underlying variable, like a propensity, that is latent
- Treats binary variables as a measurement problem

Suppose: $y_i^* = x_i\beta + \epsilon_i$

Assume: ϵ has mean 0

ϵ follows standard logistic distribution with variance $\frac{\pi^2}{3}$

Then: $y_i = 1$ if $y_i^* > t$ where t is some threshold, usually assumed = 0 for convenience
$y_i = 0$ if $y_i^* \leq t$

Logit (or probit) is just regression with less information, where all we know is the sign of y_i^*

The assumptions about the variance of ϵ and the value of t are innocuous:

You can always rescale variables. If $y^* = x\beta + \sigma\epsilon$, then you could rewrite: $\frac{y^*}{\sigma} = x\frac{\beta}{\sigma} + \epsilon$

As long as the model contains a constant, you could rewrite:

$$\Pr(y^* > a) = \Pr(\alpha + x\beta + \epsilon > a) = \Pr((\alpha - a) + x\beta + \epsilon > 0)$$

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

0. Two approaches to binary response

B. Random utility approach
- Assumes there are fixed alternatives, indexed by m
- Treats choice as if it has a systematic component (V) and a stochastic component (ϵ)

Suppose: $U_{im} = V_{im} + \epsilon_{im}$

Let: $V_{im} = X_{im}\beta_{im}$

If $m = \{1, 2\}$, then: $P(y_i = 1) = P(U_{i1} > U_{i2})$

$$= P(V_{i1} + \epsilon_{i1} > V_{i2} + \epsilon_{i2})$$

$$= P(\epsilon_{i2} - \epsilon_{i1} < V_{i1} - V_{i2})$$

Assume: ϵ is i.i.d. according to the "Type I Extreme Value distribution" $F(\epsilon_{im}) = \exp(-\exp^{-\epsilon_{im}})$

then the difference $\epsilon_{i2} - \epsilon_{i1}$ follows the standard logistic distribution

Model can easily be extended to more than two options; Daniel McFadden won the 2000 Nobel Prize for his pioneering contributions to discrete choice theory

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

1. Logit vs. Scobit (skewed logit)

logit: $P_i = \frac{1}{1+e^{-Z_i}}$ where $Z_i = \beta \mathbf{x}_i + u_i$

scobit: $P_i = \frac{1}{\left(1+e^{-Z_i}\right)^{\alpha}}$ where $\alpha > 0$; if $\alpha = 1$ then the model is logit

In logit (or in scobit when $\alpha = 1$), the maximal impact of Δx on y is at $p = 0.5$

Achen (2002) shows that the point at which the impact of Δx on y is maximal is:

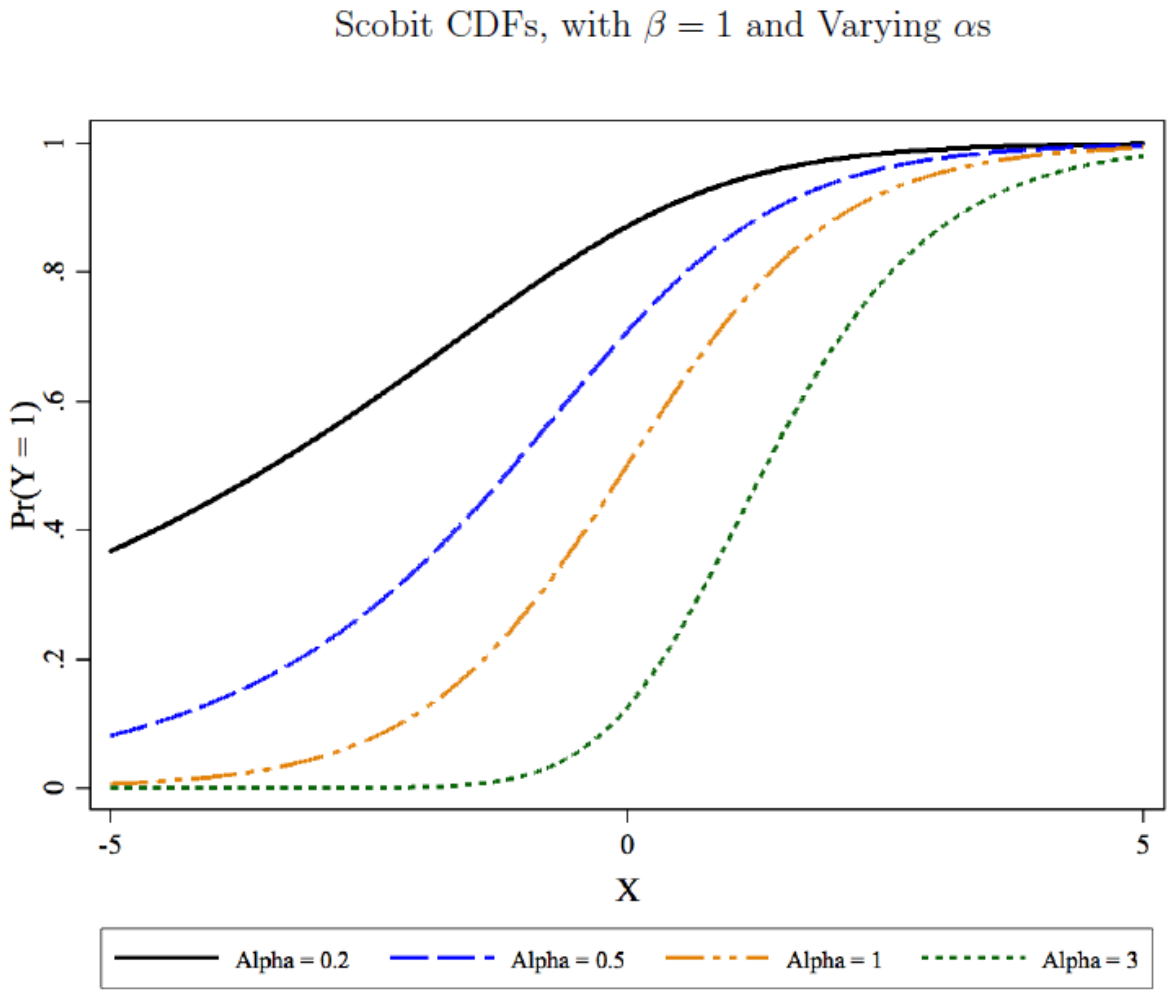
$$P^* = 1 - \left[\frac{\alpha}{\alpha+1}\right]^{\alpha}$$

In scobit when α diverges from 1, the maximal impact of Δx on y is at $p < 0.5$ when $\alpha < 1$
is at $p > 0.5$ when $\alpha > 1$

α can be though of as a shape parameter that can be estimated along with the β

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

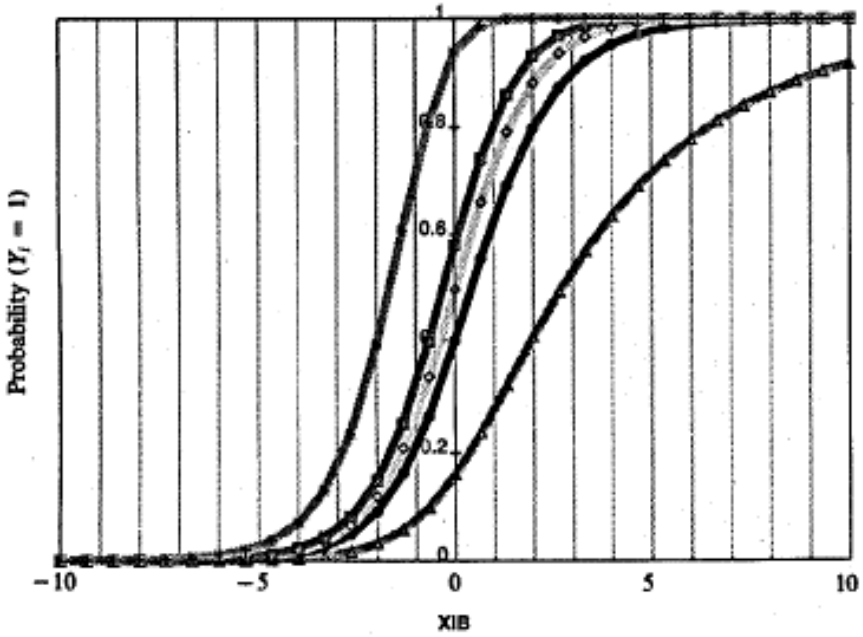
1. Logit vs. Scobit (skewed logit)



POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

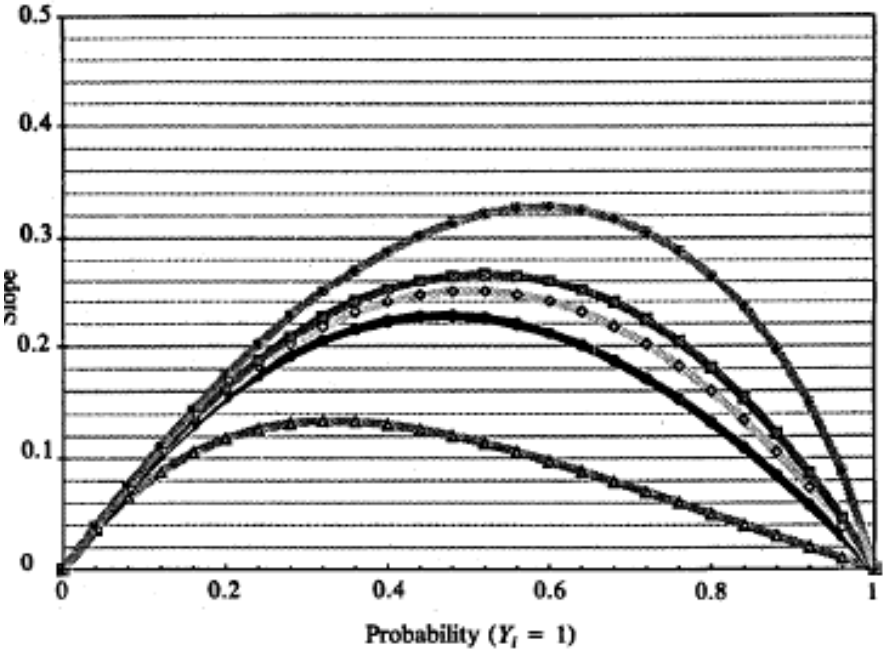
1. Logit vs. Scobit (skewed logit)

Cumulative distribution



Key:
▲α = .25
●α = .75
◇α = 1 (logit)
■α = 1.25
✱α = 4

Slope of cumulative distribution



Key:
▲α = .25
●α = .75
◇α = 1 (logit)
■α = 1.25
✱α = 4

2. Plotting Logits and Scobits

See R script: displaying logit and scobit.R

Also see R script: Default.R, which uses dataset default.csv / default01.dta

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

3. Example: *The Effect of Registration Laws and Education on U.S. Voter Turnout*

Wolfinger and Rosenstone (1980) *Who Votes?* argued that voting law restrictions have their largest impact on the least educated.

Empirical demonstration (see Nagler (1991) Table 1, p. 1397)

Education (years)	Predicted Voting Rate	Predicted Change if Day-of Registration Allowed
0-4	32.2	+ 6.2
5-7	42.2	+ 6.4
8	56.2	+ 5.8
9-11	59.0	+ 5.7
12	66.1	+ 5.2
13-15	72.1	+ 4.8
16	83.2	+ 3.6
17+	90.8	+ 2.2

Process used in *Who Votes?* described on p. 1395–1396 of Nagler (1991)

Explanation: "Formal education increases one's capacity for understanding and working with complex, abstract, and intangible subjects, … [and] this heightened level of understanding and information would also reduce the cost of registering."

However, highest education groups have highest turnout ➔ "compression" problem

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

3. Example: *The Effect of Registration Laws and Education on U.S. Voter Turnout*

Huang and Shields (2000) have a really nice visual depiction of difference between … the *Who Votes?* analysis & Nagler's (1991) reanalysis:

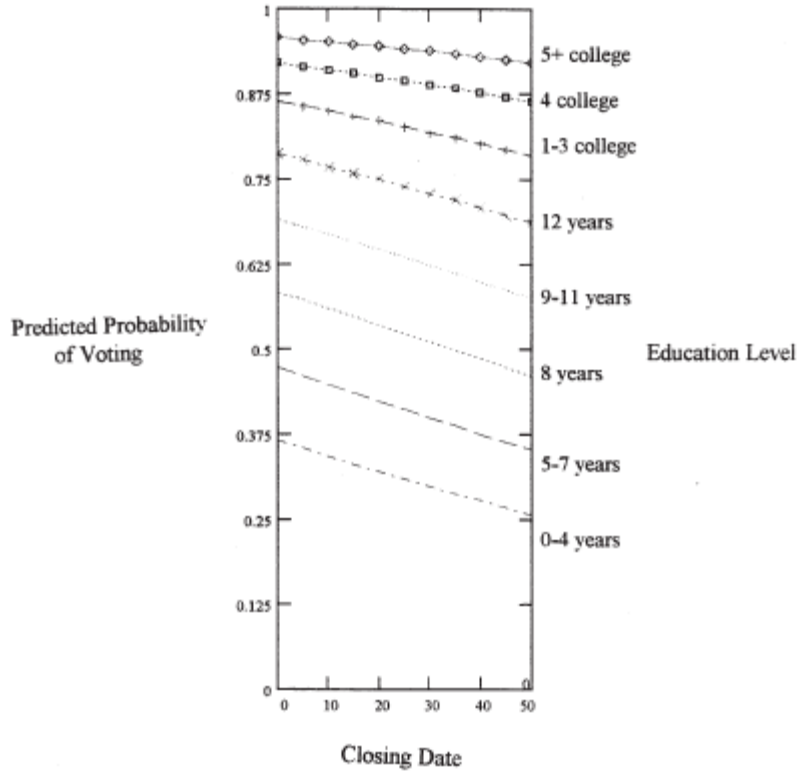


Figure 1: Predicted Probability of Voting (vertical axis) by Closing Date (horizontal axis)—Wolfinger and Rosenstone's Final Model (1972 Current Population Survey)

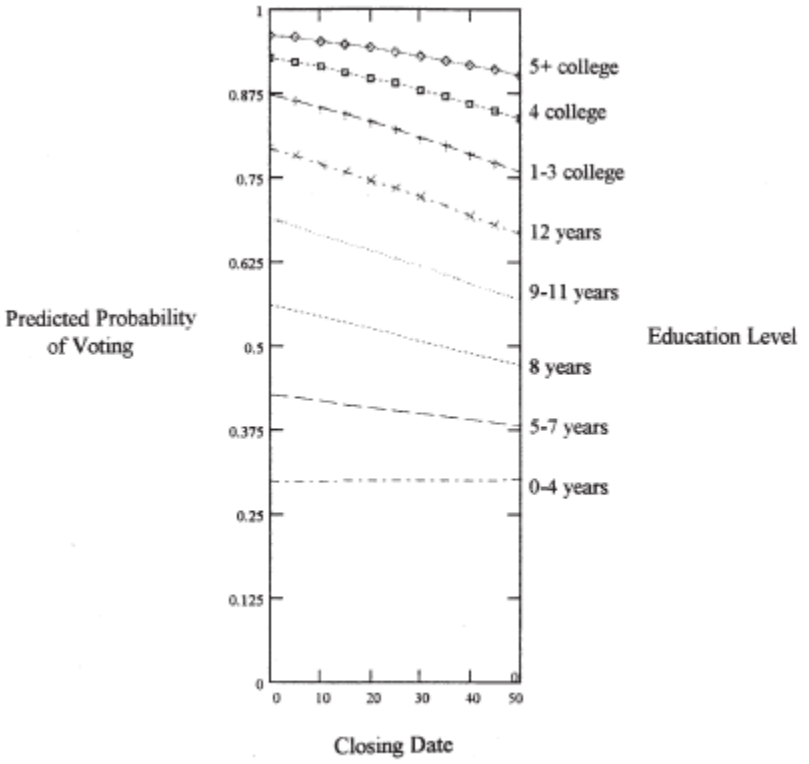


Figure 2: Predicted Probability of Voting (vertical axis) by Closing Date (horizontal axis)—Nagler's Full Model (1972 Current Population Survey)

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

3. Example: *The Effect of Registration Laws and Education on U.S. Voter Turnout*

	Replication of Wolfinger & Rosenstone			
Independent Variable	Estimated Coefficient	*t*-ratio	Estimated Coefficient	*t*-ratio
Intercept	-2.4928	-53.74**	-2.6562	-39.62**
Education	.2635	20.52**	.3003	17.83**
Education-squared	.0035	2.56**	.0033	2.36**
Age	.0653	48.69**	.0652	48.63**
Age-squared	-.0005	-35.93**	-.0005	-35.83**
South	-.1935	-15.03**	-.1936	-15.05**
Hours	—	—	—	—
Gubernatorial election	.0682	6.66**	.0683	6.67**
Irregular registration hours	-.0155	-.97	-.0137	-.86
Evening/Saturday registration	.1009	10.10**	.1009	10.10**
No absentee registration	-.0291	-1.92*	-.0295	-1.95*
Closing date	-.0062	-10.27**	-.0005	-.26
Closing date × education	—	—	-.0012	-3.38**
Closing date × education-squared	—	—	—	—
Number of cases	90,279		90,279	
Percent voting	65.30		65.30	
Correctly predicted	70.62		70.63	
Log-likelihood	-51,915		-51,909	

POLS 6481. Research Design and Quantitative Methods II
Lecture 19. Scobit Models
Readings: Nagler

3. Example: *The Effect of Registration Laws and Education on U.S. Voter Turnout*

Independent Variable	Replication of Wolfinger & Rosenstone: Estimated Coefficient	Replication of Wolfinger & Rosenstone: *t*-ratio	Estimated Coefficient	*t*-ratio
Intercept	-2.4928	-53.74**	-2.7597	-26.51**
Education	.2635	20.52**	.3544	7.90**
Education-squared	.0035	2.56**	-.0029	-.59
Age	.0653	48.69**	.0652	48.60**
Age-squared	-.0005	-35.93**	-.0005	-35.79**
South	-.1935	-15.03**	-.1939	-15.07**
Hours	—	—	—	—
Gubernatorial election	.0682	6.66**	.0686	6.70**
Irregular registration hours	-.0155	-.97	-.0134	-.84
Evening/Saturday registration	.1009	10.10**	.1010	10.11**
No absentee registration	-.0291	-1.92*	-.0293	-1.94*
Closing date	-.0062	-10.27**	.0032	.96
Closing date × education	—	—	-.0032	-2.07**
Closing date × education-squared	—	—	.0002	1.30
Number of cases	90,279		90,279	
Percent voting	65.30		65.30	
Correctly predicted	70.62		70.61	
Log-likelihood	-51,915		-51,909	

4. Example, continued: *Scobit: An Alternative Estimator to Logit*

	Logit[a]		Scobit[b]	
Independent Variable	Estimated Coef.	*t*-Stat.	Estimated Coef.	*t*-Stat.
Intercept	−4.4151	−24.60*	−4.7514	−28.13*
Education	0.3590	5.01*	0.2055	5.81*
Educ. squared	0.0193	2.61*	0.0706	10.27*
Age	0.1140	51.92*	0.1798	19.60*
Age squared	−0.0008	−36.14*	−0.0013	−19.27*
South	−0.1834	−10.55*	−0.2803	−11.00*
Gub. Elec.	0.0063	0.33	0.0019	0.09
Closing date	0.0007	0.11	−0.0213	−14.65*
Closing date × educ.	−0.0046	−1.79**	—	—
Closing date × educ-sq.	0.0004	1.32	—	—
α	—	—	0.4242	13.30*
$P^*(\alpha)$			0.40	
Number of cases	98,857		98,857	
Percent voting	67.01		67.01	
Correctly predicted	71.08		71.30	
Log-likelihood	−55,331[c]		−55,289	