*Scobit: An Alternative Estimator to Logit and Probit**

Jonathan Nagler, *University of California, Riverside*

Logit and probit, the two most common techniques for estimation of models with a dichotomous dependent variable, impose the assumption that individuals with a probability of .5 of choosing either of two alternatives are most sensitive to changes in independent variables. This assumption is imposed by the estimation technique because both the logistic and normal density functions are symmetric about zero. Rather than let methodology dictate substantive assumptions, I propose an alternative distribution for the disturbances to the normal or logistic distribution. The resulting estimator developed here, scobit (or skewed-logit), is shown to be appropriate where individuals with *any* initial probability of choosing either of two alternatives are most sensitive to changes in independent variables. I then demonstrate that voters with initial probability of voting of *less than* .5 are most sensitive to changes in independent variables. And I examine whether individuals with low levels of education or high levels of education are most sensitive to changes in voting laws with respect to their probability of voting.

Introduction

Nonlinear models such as logit and probit have gained favor among political scientists as ways to overcome the efficiency and specification problems of ordinary least squares (OLS) when estimating models with dichotomous dependent variables. Two features are inherent in such models. First, the effects of changes in independent variables depend upon the initial value of the dependent variable (i.e., of the probability that the dependent variable takes on each value). Second, such models are "interactive" in all of their variables: the effect of a change in any independent variable upon the dependent variable will depend upon the values of all of the other independent variables. These properties suggest that care should be taken in the discovery of systematic differences in sensitivity to stimuli across respondents (henceforth referred to as respondent heterogeneity), or interactive effects between variables in the model, since both these phenomena are assumed—and in fact *imposed*—by the model specification.

*I would like to thank John Freeman, Simon Jackman, John Jackson, and reviewers of this journal for helpful comments. Earlier versions of this paper were presented at the annual meeting of the Political Methodology Society, Cambridge, July 1992, and the annual meeting of the Midwest Political Science Association, Chicago, April 1992.

American Journal of Political Science, Vol. 38, No. 1, February 1994, Pp. 230–55

In particular, probit and logit will tend to exaggerate effects of changes in any independent variables for those individuals having a probability closest to one-half of choosing either of the two alternatives (i.e., for those individuals with P_i = prob(Y_i = 1) = .5). Imagine that persons with a 50% likelihood of voting share a common trait: they are poorly educated. Then when we examine changes in individuals' likelihood of voting caused by changes in some other explanatory variable, we shall see an exaggerated change for poorly educated persons. Thus, we might conclude that poorly educated persons are most sensitive to stimuli, when in fact this is an assumption of the model. Such an observation has been interpreted to indicate that poorly educated individuals have special problems dealing with voting laws (Wolfinger and Rosenstone 1980; for a revision to this claim, see Nagler 1991).

Alternatively, imagine that one wishes to examine the effect of a congressional challenger's campaign spending on the likelihood of the challenger's partisans switching allegiance during the campaign versus the effect of the challenger's spending on the likelihood of the incumbents' partisans switching allegiance during the campaign. Any estimates of predicted change would be "contaminated" by the starting points of the probabilities of switching (Jacobson 1990). Or in another campaign setting, imagine testing whether campaign canvassing has larger effects on poor persons or on rich persons. Presumably such questions are of considerable importance to a campaign organization deciding how to allocate its resources strategically. Yet probit or logit estimates would confront the same methodological problem in this case: any estimates of change would depend upon initial probabilities of the respondent choosing either option. And since the probability of choosing either option is likely to be correlated with the individual characteristic of interest (partisanship or income), then probit or logit are *assuming* an interactive effect between the campaign activity and the individual characteristic.

In linear models of the form $Y_i = X_i\beta$, there are two common techniques to test for interactive effects. First, a multiplicative interactive term can be added between the individual characteristic of interest and the independent variable of interest. Or, second, the data can be disaggregated by the characteristic of interest and the coefficients of the variable of interest can be compared across samples. Such tests are dependent upon the fact that the marginal effect of x_k upon Y ($\partial Y/\partial x_k$), the change in the dependent variable caused by a change in the independent variable, is determined solely by β_k, the coefficient of interest. However, in nonlinear models, this is not the case: $\partial P_i/\partial x_k$ is dependent upon both β_k and $f(X_i\beta)$, where f is the density function assumed for the disturbances. If the density function assumed is wrong, then estimates of marginal effects and interactive effects will be wrong. Below I develop the standard framework for the binary response model, then show that by choosing a set of distributions for the disturbances dependent upon a parameter to be estimated, rather than assuming a specific distribution, it is possible to estimate the correct specification and hence correctly estimate marginal and interactive effects.

The Binary Response Model

Following the usual procedure when dealing with dichotomous variables, assume that while we observe only the values of zero and one for the variable Y, there is a latent, unobserved continuous variable Y^* that determines the value of Y.[1] Furthermore, assume that Y^* can be specified as follows:

$$Y_i^* = X_i\beta + u_i, \tag{1}$$

and that

$$\begin{aligned} Y_i &= 1 \quad \text{if} \quad Y_i^* > 0 \\ Y_i &= 0 \quad \text{otherwise,} \end{aligned}$$

where X represents a vector of random variables, and u represents a random disturbance term. Now from equation (1):

$$P_i = \text{prob}(X_i\beta + u_i > 0). \tag{2}$$

Rearranging terms,

$$P_i = \text{prob}(u_i > (-X_i\beta)) \tag{3}$$

$$= 1 - F(-X_i\beta), \tag{4}$$

where F is the cumulative density function of the variable u.

Now the marginal effect on P_i for a change in X_k is given by:

$$\begin{aligned} \frac{\partial P_i}{\partial(X_k)} &= \frac{\partial[1 - (F(-X_i\beta))]}{\partial(X_k)} \\ &= f(-X_i\beta)\beta_k \end{aligned} \tag{5}$$

Thus, the impact of changes in a variable X_k on the likelihood of a particular individual choosing option number 1 will depend not only on β_k (the variable's coefficient), but also on the value of $X_i\beta$, and in particular $f(-X_i\beta)$. Since $\partial P_i/\partial(X_k)$ will depend upon the choice of F, the true F must be known in order to know the true impact of changes in any inde-

[1] See Maddala (1983) for a thorough treatment of limited dependent variables.

pendent variable upon different individuals. Or, the shape of the true $F(u)$, and $f(u)$, will depend upon which individuals are most sensitive to changes in the independent variables.

If we assume that u is normally distributed as in the probit model, or that $F(u) = \Phi(u)$, then $f(-X_i\beta) = \phi(-X_i\beta)$, and $f(-X_i\beta)$ has a maximum at $X_i\beta = 0$. This is precisely where $\Phi(X_i\beta) = .5$, and hence $P_i = .5$. This implies that any given variable X_k will have its greatest effect on those individuals for which $X_i\beta$ is closest to 0, or for which P_i is closest to 0.5. Or if the previous statement about the sensitivity of P_i is correct, then $F(u) = \Phi(u)$ is the correct distribution and probit is the appropriate estimation technique. However, if individuals with initial probability other than .5 are those most sensitive to change, then the probit model would represent a misspecification and lead to biased inferences about the marginal effect of changes of any independent variable. In particular, it would invalidate inferences that certain individuals are more sensitive to stimuli based solely on predicted probabilities derived from probit estimates. And since logit is also based on a symmetric distribution, the same criticism would apply to it as well.

We would like to move beyond the world described by Poincaré in which "everyone believes the [normal] law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an empirical fact" (cf. Harvey 1981).[2] The goal of this research is to specify correctly the response curve so as to determine precisely which individuals are most sensitive to change. Since we can distinguish between individuals based on their initial probability to choose an option (vote or not vote, for instance), this is equivalent to finding the value of P_i where sensitivity—$\partial P_i/\partial(X_k)$—is at a maximum. In other words, the goal of this research is to allow the data to suggest the most reasonable response curve, rather than assuming that the logit or probit curves fit best. This is accomplished via maximum likelihood estimation of an additional parameter, α, that modifies the response curve so that the probability level at which independent variables have maximum impact on change in probability is not necessarily .5, but is instead determined by the actual patterns observed in the data.

The Estimation Technique

The problem of determining an appropriate response curve is simplified if we consider $f(z)$, where $z_i = X_i\beta$, X_i is a vector of k independent variables, and β is a vector of k parameters. For a given distribution the question of interest is, at what value P^* of P_i is $\partial P_i/\partial z$ at a maximum?

[2]I found this wonderfully appropriate quote in King (1989).

Or for a given distribution, at what probability level are individuals most sensitive to stimuli? We would like to test a set of distributions, each differing in P^*, where $z_{\max[f]} = F^{-1}(P^*)$ maximizes $f(z)$. And we would like to find an estimation technique to allow us to determine the correct distribution from among this set. In other words, we would like to find a set of distributions $\{F_1, F_2, F_3, \ldots ,\}$ such that for every value of P^* in the interval (0, 1) there exists some F_j such that $z = F_j^{-1}(P^*)$ maximizes $f_j(z)$. These distributions should be well behaved over the range $-\infty < X_i\beta < \infty$. The distribution within that set fitting the data best (i.e., the distribution most likely to generate the observed data by maximum likelihood criteria) would be our choice as the true distribution. Obviously insisting only that the set of distributions satisfy the above criteria does not cover all possible distributions. However, it would give us a distribution corresponding to any possible set of individuals—based on their initial probability of choosing either alternative—being most sensitive to changes in the independent variables. The alternative would be to adopt a semiparametric estimation technique that would not assume any functional form for the distribution (Härdle 1990). If data-gathering technology increased at the same speed as computing technology increased, this would be feasible. However, it is unlikely that the typical political science data set is up to the task of allowing for precise estimation this way.

Now, by adding a parameter to the definition of the distribution, we may attempt to describe a set of distributions with the above criteria. The following distribution, one of several proposed by Burr (1942) and referred to here as the Burr-10 distribution to distinguish it from the more commonly used distribution associated with Burr's name, is adopted:[3]

$$F(z; \alpha) = \frac{1}{(1 + e^{-z})^{\alpha}}, \tag{6}$$

where $\alpha > 0$. The Burr-10 distribution satisfies the condition that $f(z)$ not attain a maximum only when $F(z) = .5$, *and* it is defined for $-\infty < z < \infty$. It remains to be shown that the Burr-10 distribution meets the criteria set forth (i.e., that $\forall\, P^* \in (0, 1)\ \exists\ \alpha$ s.t. $z = F_j^{-1}(P^*)$ maximizes $f_j(z; \alpha)$). The proof of the following proposition (Appendix A) shows that this condition can be met by this set of distributions.

PROPOSITION 1: For all $P^* \in (0, 1)$, there exists $\alpha > 0$, s.t. $z = F_j^{-1}(P^*)$ maximizes $f_j(z; \alpha)$, where $F(z; \alpha) = (1 + e^{-z})^{-\alpha}$.

[3]This distribution was given in equation (10) of Burr's (1942) work.

The Burr-10 distribution may thus be used to generate an alternative estimator to probit or logit. This estimator is called the *scobit* estimator here, or "skewed-logit" because it allows for a skewed response curve, with α serving as a parameter to measure skewness.[4]

Other more common distributions fail this property (i.e., that $\forall\, k \in (0, 1)\ \exists\, \alpha$ s.t. $P^* = k$). For the normal distribution, logistic distribution, and extreme value distribution, it can be shown that $f(z)$ can be defined to within a constant scale factor as functions of $F(z)$. Hence, for these distributions, P^* is invariant with respect to the choice of parameters. For example, the density function for the extreme value distribution can be expressed as a function of its distribution as follows:

$$f(z) = (-1/\beta)\log(F(z))F(z). \qquad (7)$$

This means that $z_{\max[f]}$ will correspond to a unique $F(z)$. By contrast, for the Burr-10 distribution, we have:

$$f(z) = \alpha(F(z)^{-1/\alpha} - 1)(F(z))(F(z)^{1/\alpha}). \qquad (8)$$

Hence, depending upon the value of the parameter α, $z_{\max[f]}$ could correspond to many (any, as I have shown) values of $F(z)$.

With the addition of new parameters, some distributions could be modified to offer the property in question. However, the Burr-10 distribution is also desirable because the logistic distribution is nested within it. If we adopt the constraint that $\alpha = 1$, then the Burr-10 distribution is the logistic distribution. This nesting property allows for log-likelihood ratio tests comparing scobit to logit; LL_{logit}/LL_{scobit} will have a χ^2 distribution with one degree of freedom. Hence, throughout the paper, scobit estimates are compared to logit rather than probit.[5] This nesting property is shared with an estimator proposed by Prentice (1976) for use in estimating dose response curves. However, Prentice's estimator requires fitting two additional parameters to the data to describe the distribution, rather than only one. Thus, using the Burr-10 distribution allows for a more elegant test, since the estimate of only one parameter need be evaluated.

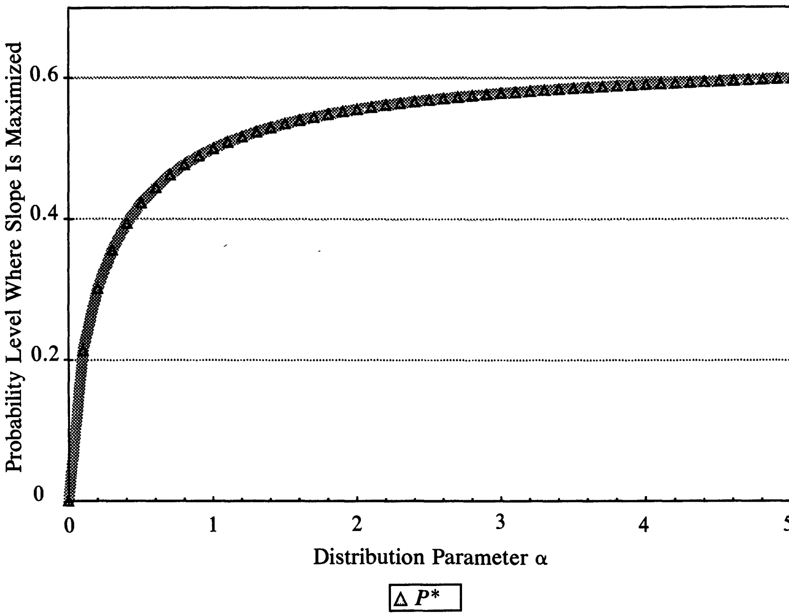
The Scobit Estimator

Figure 1 illustrates the relationship between the parameter α and the value of P^* that maximizes $\dot{P}_i$ or that maximizes sensitivity of the

[4]The naming and spelling of "scobit" are based on the preferences of the author; "skewed logit" would be more in keeping with the conventions of the econometrics literature.

[5]See Aldrich and Nelson (1984, 42) for a comparison of logit and probit estimates.

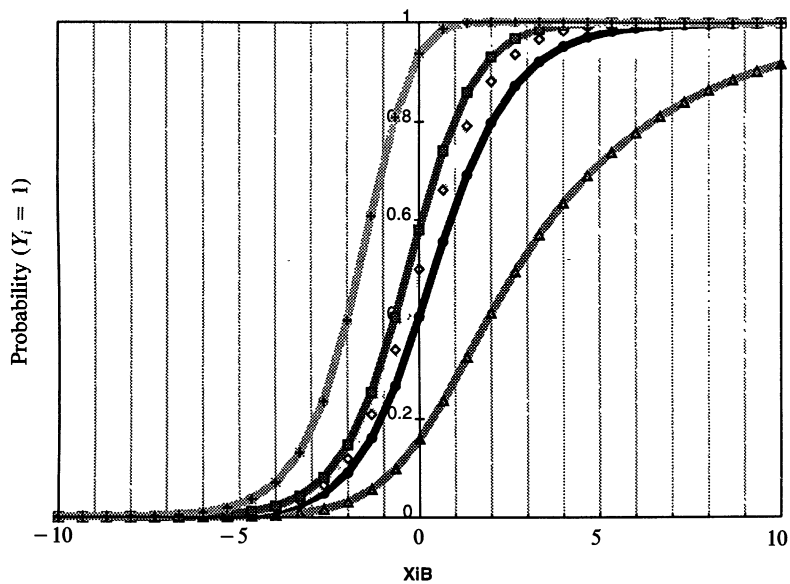
Figure 1. Probability Level Where Slope Is Maximized for Different Values of the Parameter α



respondent to changes in the independent variables. For $\alpha = .1$, $\dot{P}_i$ achieves a maximum where $P_i \cong .2$; for $\alpha = .2$, $\dot{P}_i$ achieves a maximum where $P_i \cong .3$. The curve flattens out as α increases, indicating that P^* reaches a limit as $\alpha \rightarrow \infty$. The figure shows that this limit is approached quite rapidly. Thus, if individuals with high initial probabilities of choosing alternative 1 are most sensitive to stimulus, the parameter value α should be high. If individuals with low initial probabilities of choosing alternative 1 are most sensitive to stimulus, then the parameter value of α should be low. In fact, for $\alpha = 1$, $P^* = .5$, and $f(-z)$ is symmetric about this. For values of α less than 1, respondents unlikely to choose alternative 1 are most sensitive to stimuli. For values of α greater than 1, respondents likely to choose alternative 1 are most sensitive to stimuli.

Figure 2.A compares the cumulative distribution for logit to the cumulative distribution for scobit for values of $\alpha = .25$, .75, 1.00 (logit), 1.25, and 4.00. Figure 2.B plots the density (slope) for scobit for the same values of α against the associated probability value. As can be seen, the slope takes on its maximum values at different probability levels depending upon the value of α. Figure 2.A gives a good indication of how

Figure 2.A. Cumulative Distribution for Scobit
[Y − Axis = 1 − $F(-X_i\beta)$]



Key:

▲α = .25

●α = .75

◇α = 1 (logit)

■α = 1.25

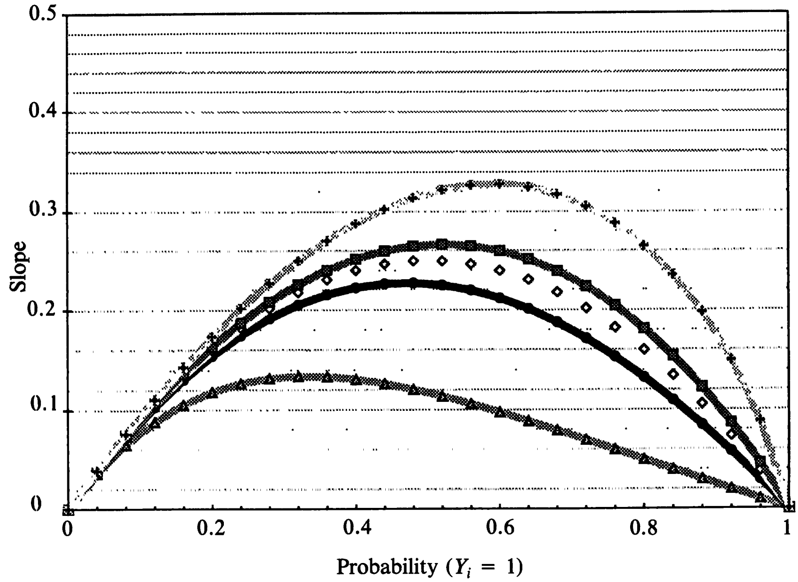
✱α = 4

scobit works compared to logit. As the parameter α varies, the curve is not uniformly shifted, but rather the shape of the curve changes: the point where it is steepest varies. Figure 2.B indicates more directly the relationship between P_i and $\partial P_i/\partial Z$. The apogee of each curve is the point where $\partial P_i/\partial Z$ is at a maximum. The figure clearly shows how increasing the value of α increases the probability level at which this apogee occurs.

Scobit versus Logit: Monte Carlo Results

Before proceeding to analyze actual data, I present Monte Carlo experiments to illustrate: (1) how well scobit will behave in small samples where the disturbances in the model are both symmetrically and asymmetrically distributed and (2) how much scobit estimates will differ from

Figure 2.B. Slope for Scobit at Different Values of α versus Probability $Y = 1$



Key:

▲α = .25

●α = .75

◇α = 1 (logit)

■α = 1.25

✱α = 4

logit estimates when the disturbances are asymmetrically distributed. Again, scobit estimates are compared to logit rather than probit because logit is actually a constrained version of scobit: the constraint being that α is fixed at the value 1.00.

Table 1 reports the results of 100 trials with both 500 and 2,000 observations per trial where the disturbance in the true model was logistically distributed. Variables Y^* and Y were generated as follows:

$$Y_i^* = -10 + 2X_i + u_i,$$
$$Y = 1 \quad \text{if} \quad Y^* > 0$$
$$Y = 0 \quad \text{if} \quad Y^* \leq 0,$$

where u_i was a random variable with logistic distribution, and X was a random variable uniformly distributed between 0 and 10 so that $E[Y^*]$

Table 1. Monte Carlo Experiment No. 1
(Results of 100 Trials with Logistically Distributed Disturbance)

		Logit		Scobit	
	True	Mean	Std. Dev.	Mean	Std. Dev.
N = 500					
β_0	−10.00	−10.28	1.39	−11.18	3.63
β_1	2.00	2.06	0.27	2.25	0.98
α	1.00	—	—	1.48	1.32
LLF	—	−78.91	—	−78.34	—
CorPred[a]		466.57	5.10	466.91	5.08
N = 2,000					
β_0	−10.00	−10.01	0.59	−10.03	0.77
β_1	2.00	2.00	0.12	1.99	0.22
α	1.00	—	—	1.10	0.31
LLF	—	−329.97	—	−329.50	—
CorPred		1,860.20	10.68	1,860.50	10.84

[a]CorPred is the number of correct predictions.

= 0 and $E[Y]$ = .5. Logit and scobit estimates were computed for the resulting data. The mean value of the estimate of the scobit skewness parameter α over the 100 trials of sample size 500 was 1.48 (σ_α = 1.32) and was 1.10 for the 100 trials of sample size 2,000 (σ_α = 0.31). Thus, with 2,000 observations, the skewness parameter of scobit was able to cluster fairly tightly about the true value of 1.00; with only 500 observations, the estimate of skewness was very imprecise.[6]

The larger standard errors of the estimated scobit coefficients (β's) versus the estimated logit coefficients suggest a price to pay in using scobit. However, the scobit estimates are not intractable. Even with only 500 observations, the β's are significantly different from 0, if not estimated very precisely. The log-likelihood values and number of correct predictions were virtually indistinguishable for both logit and scobit; the average log-likelihood value was within .6 for both experiments. The average scobit log-likelihood value was slightly lower, as would have to be the case since scobit removes a constraint from the logit model.

[6]One possible reason for the difference between the estimated and true parameters in the trials of sample size 500 is that some parameter estimates may simply be wrong; the software may have converged on the wrong point. This would really have to be hand checked for the extreme estimates.

The next question is, How does logit do versus scobit when the disturbance term is *not* symmetric (i.e., what happens when scobit rather than logit represents the true underlying model)? Table 2 reports the results of 100 trials with both 500 and 2,000 observations per trial where the disturbance in the true model was distributed as a Burr-10 random variable, with $\alpha = .5$. This corresponds to a case where $P^* = .42$. Variables Y^* and Y were generated as follows:

$$Y_i^* = -8.5 + 2X_i + u_i,$$
$$Y = 1 \quad \text{if} \quad Y^* > 0$$
$$Y = 0 \quad \text{if} \quad Y^* \leq 0,$$

where u_i was a random variable with Burr-10 distribution; $\alpha = .5$. Table 3 was generated in a similar manner with $\alpha = .25$ ($P^* = .33$), and coefficients of -6.50 and 2.00. The coefficients are changed from the first example to retain an equal distribution of the observed variable Y; with these parameters $E[Y] = .5$.

In Table 2 the average estimate of the scobit parameter α was .62 for the experiment with sample size 500. Since the standard error was .38, this suggests that with such a small sample, we would not be able to reject the null hypothesis that $\alpha = 1$ (i.e., that logit is the correct model)

Table 2. Monte Carlo Experiment No. 2 (Results of 100 Trials with Burr-10 Distributed Disturbance [$\alpha = .5$])

		Logit		Scobit	
	True	Mean	Std. Dev.	Mean	Std. Dev.
$N = 500$					
β_0	−8.50	−7.14	0.78	−9.37	3.20
β_1	2.00	1.44	0.16	2.20	0.96
α	0.50	—	—	0.62	0.38
LLF	—	−116.07	—	−114.47	—
CorPred[a]		450.88	6.59	451.56	6.22
$N = 2,000$					
β_0	−8.50	−7.02	0.33	−8.56	0.76
β_1	2.00	1.42	0.07	2.02	0.23
α	0.50	—	—	0.51	0.10
LLF	—	−461.68	—	−456.05	—
CorPred		1,805.90	11.43	1,808.10	11.80

[a]CorPred is the number of correct predictions.

Table 3. Monte Carlo Experiment No. 3
(Results of 100 Trials with Burr-10 Distributed Disturbance [α = .25])

	True	Logit		Scobit	
		Mean	Std. Dev.	Mean	Std. Dev.
N = 500					
β_0	−6.50	−4.54	0.39	−9.20	5.69
β_1	2.00	0.90	0.08	3.00	2.17
α	0.25	—	—	0.23	0.11
LLF	—	−172.72	—	−166.03	—
CorPred[a]		425.45	7.61	427.82	7.84
N = 2,000					
β_0	−6.50	−4.53	0.16	−6.63	0.87
β_1	2.00	0.90	0.03	2.04	0.37
α	0.25	—	—	0.26	0.05
LLF	—	−698.24	—	−677.75	—
CorPred		1,694.80	13.59	1,703.20	12.44

[a]CorPred is the number of correct predictions.

at traditional significance levels. In the trials with a sample size of 2,000, the estimate of α was .51, and the standard error dropped to .10; thus rejection of α = 1 is possible at very high levels of significance ($\rho < .01$). According to Table 3, with α = .25, it is possible to reject the hypothesis that α = 1 at traditional levels in samples of only 500: the average estimated value of α is 0.23, with $\sigma_\alpha = .11$. With sample size 2,000, the estimate of α was even more precise. Thus, under these scenarios, rejection of logit as the correct model is possible.

Estimates of β and Predicted Probabilities

With both sample sizes of 500 and 2,000, scobit generated coefficient estimates for the β's that are significant at traditional levels for α = .50. With α = .25, the coefficients for sample size 500 were not significant at traditional levels. The logit coefficients (β's) were also statistically significant for α = .50 and α = .25 at both sample sizes. However, *they are wrong:* when the assumed underlying model (logistic disturbances) is wrong, then the maximum likelihood estimates are not consistent, and we cannot make any valid statistical statements about them. So the question becomes, does it matter that logit is used when it is not the correct model? While α = 1 could be rejected at the 99% level of significance, the substantive findings from the scobit estimates may not differ very much from those obtained with logit (i.e., with $\alpha = 1$). The mean number of correct predictions out of the sample of 2,000 goes from 1,805.9 to 1,808.1 when using scobit instead of logit for $\alpha = .50$; and it rises from 1,694.8 to 1,703.2 for $\alpha = .25$.

While neither of the above differences may be perceived as overwhelming, the relevant consideration if we are interested in testing for heterogeneity of respondents, or interactive effects, is the *pattern* of correct predictions. If logit were to overpredict for individuals with "low values" of X and underpredict for individuals with "high values" of X, then scobit would be much preferred to it. This of course is testable. The Monte Carlo experiments provide coefficient estimates that logit produces even though, since the disturbances are *not* distributed logistically, these coefficients do not represent estimates of the true underlying model. Nonetheless, these coefficients can be used to indicate what predicted probabilities logit would produce for different values of X. The predicted logit probabilities are denoted by $\hat{\hat{P}}$, where $\hat{\hat{P}} = \text{prob}(Y_i = 1 | X = X_i)$; $\hat{\hat{P}}$ was computed as $1 - F_L(-X_i\hat{\hat{\beta}})$, where $\hat{\hat{\beta}}$ was the average logit estimate produced by the Monte Carlo experiments and F_L is the cumulative logistic distribution. The predicted scobit probabilities are denoted by $\hat{P}$, which was computed as $1 - F_S(-X_i\beta; \alpha)$, where F_S is the Burr-10 distribution used in the scobit estimator.

Table 4 illustrates the comparison between the logit and scobit predicted probabilities for three values of α: .50, .25, and .10, and values of X ranging from 0 to 10 in increments of 0.5. For $\alpha = .5$, the predicted probabilities for the two specifications are never too far apart. The largest gap is reported at $X = 5$, where $\hat{\hat{P}} = .52$ and $\hat{P} = .57$. For $\alpha = .25$, the differences are greater. The largest gap reported here is again for $X = 5$, where $\hat{\hat{P}} = .49$ and $\hat{P} = .59$. Finally, for $\alpha = .10$ ($P^* = .21$), there are very large differences at low values of P, $\hat{\hat{P}} = .11$ when $\hat{P} = .02$, and smaller differences for $\hat{P}$ closer to .50. Figure 3.A–3.C illustrates this by plotting the values of $\hat{\hat{P}}$ and $\hat{P}$ for each of the three values of α against X. The differences between the curves indicates the errors in the logit predictions. For all three values of α, the curves cross twice (offer the same prediction) and have larger gaps near $P = .5$ and toward the two tails. Thus, logit underpredicts in the middle and overpredicts at the tails in these cases.

Summary of Monte Carlo Results

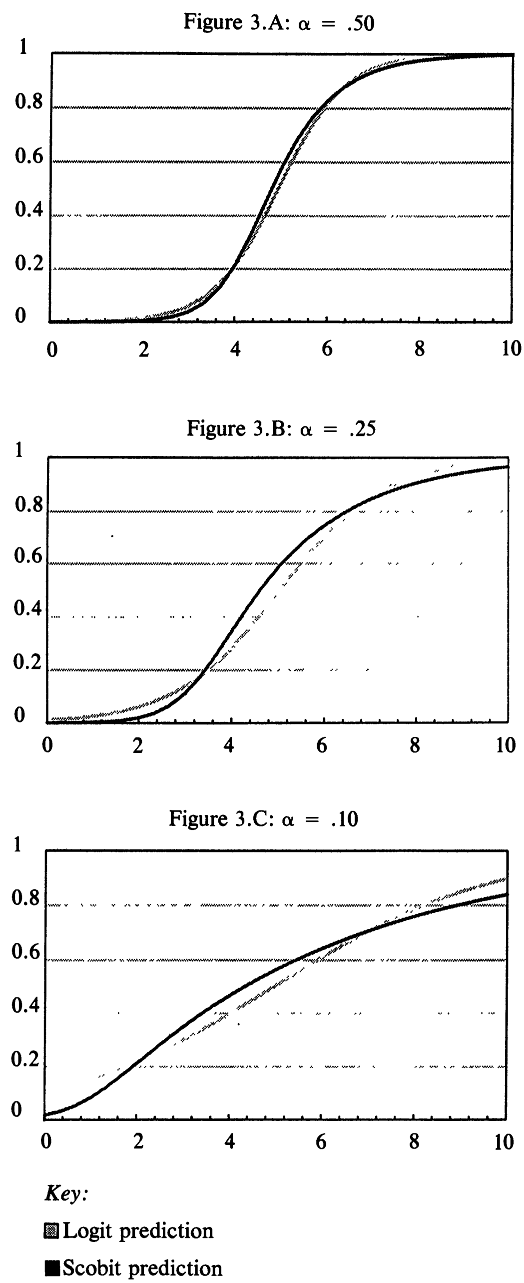
Thus, when the disturbances are logistically distributed, as the logit model assumes, scobit produces consistent, but inefficient, estimates. The relevant questions would be how much the scobit predictions differ

Table 4. Predictions Based on Monte Carlo Estimates (Logit vs. Scobit Probability Estimates)

	$\alpha = .50$		$\alpha = .25$		$\alpha = .10$	
X	$\hat{P}$ Logit	$\hat{P}$ Scobit	$\hat{P}$ Logit	$\hat{P}$ Scobit	$\hat{P}$ Logit	$\hat{P}$ Scobit
.00	.00	.00	.01	.00	.11	.02
.50	.00	.00	.02	.00	.13	.04
1.00	.00	.00	.03	.00	.15	.08
1.50	.01	.00	.04	.01	.18	.14
2.00	.02	.01	.06	.02	.22	.21
2.50	.03	.01	.09	.05	.26	.28
3.00	.06	.04	.14	.11	.30	.35
3.50	.11	.10	.20	.22	.35	.41
4.00	.21	.21	.28	.35	.40	.47
4.50	.35	.39	.38	.48	.45	.52
5.00	.52	.57	.49	.59	.50	.56
5.50	.69	.72	.60	.68	.56	.61
6.00	.82	.83	.70	.75	.61	.64
6.50	.90	.90	.79	.80	.66	.68
7.00	.95	.94	.85	.85	.71	.71
7.50	.97	.96	.90	.88	.75	.74
8.00	.99	.98	.94	.93	.82	.78
9.00	1.00	.99	.97	.94	.85	.80
9.50	1.00	.99	.98	.96	.88	.82
10.00	1.00	1.00	.99	.97	.90	.84

from the logit predictions for different levels of skewness, as well as how precisely it is possible to estimate the parameter α. If we could never be sure that $\alpha \neq 1$, then there is little to gain by running scobit rather than logit. However, if we can reject the null hypothesis that $\alpha = 1$ at an appropriate level of significance using available data (i.e., if such a rejection does not require unrealistically large samples), then the scobit model may be preferred in such cases to the logit model. This might be the case *even* if the logit model yielded a comparable log-likelihood statistic; it would depend upon the substantive question of interest. In practice, one could run scobit; and if one could not reject with some high level of confidence the hypothesis that $\alpha \neq 1$, then one would probably prefer to use the logit estimates. The Monte Carlo results presented here suggest that scobit may perform acceptably well for small samples and that logit and scobit estimates will differ.

Figure 3.A–C. Scobit Predictions versus Logit Predictions



A Test of Substantive Implications of Different Distributions

I next test the scobit estimator against logit on a model of voting turnout. As noted above, an earlier conclusion reached regarding the effect of restrictive voting laws on turnout was that such laws were particularly onerous for poorly educated individuals. Using data from the 1984 Census Bureau Current Population Survey (CPS), I reestimate the model initially estimated by Wolfinger and Rosenstone (1980), as well as the model as modified by Nagler (1991). The model postulates that an individual's probability of voting is a function of the individual's age and education, as well as the region the individual lives in (South/Non-South), whether or not there is a gubernatorial election going on in the state, and the number of days before the election that registration closes (*closing*). Nagler modified this model to include a multiplicative interactive term for *education* × *closing*. The point of this research is to determine whether poorly educated individuals have particular problems with the closing requirement and to determine which individuals are most sensitive to stimuli of any sort.

Using the procedure described above of parameterizing the distribution, I test two hypotheses. First, the estimate of the skewness parameter α reveals which voters are most sensitive to stimulus: those initially likely to vote or those initially likely to stay home. Second, estimates of P_i based upon the correct distribution of the disturbances allow calculations of $\Delta P/\Delta closing$ for individuals with different levels of education; thus the substantive question of whether poorly educated individuals or highly educated individuals are more affected by changes in voting laws can be answered, as well as the question of how much of this has to do with sensitivity peculiar to changes in voting law versus changes in *any* factors likely to increase turnout.

Some Empirical Results

The results of probit, logit, and scobit estimates of the full specification, with interactive terms explicitly added for education and closing, are reported in columns 1–6 of Table 5. As can be seen by comparing the log-likelihood statistics, the scobit model outperforms the probit and logit models. According to log-likelihood ratio tests, the scobit model is preferred to the logit model at the 99% level of confidence. The percentage correctly predicted also goes up from 71.08 with the logit estimates and 71.04 with the probit estimates, to 71.30 with the scobit estimates. While the improvement in fit may be less than overwhelming, the goal is to improve estimates of effects more so than to improve overall fit.

The estimated value of the skewness parameter α is .42; thus we can

Table 5. Probit, Logit, and Scobit Estimates of a Model of Voter Turnout, 1984

	Probit[a]		Logit[a]		Scobit[a]		Scobit[b]	
Independent Variable	Estimated Coef.	*t*-Stat.	Estimated Coef.	*t*-Stat.	Estimated Coef.	*t*-Stat.	Estimated Coef.	*t*-Stat.
Intercept	−2.7443	−25.61*	−4.4151	−24.60*	−5.3290	−17.37*	−4.7514	−28.13*
Education	0.2647	6.39*	0.3590	5.01*	0.3516	3.28*	0.2055	5.81*
Educ. squared	0.0051	1.23	0.0193	2.61*	0.0654	4.86*	0.0706	10.27*
Age	0.0696	53.08*	0.1140	51.92*	0.1822	18.50*	0.1798	19.60*
Age squared	−0.0005	−37.44*	−0.0008	−36.14*	−0.0013	−18.13*	−0.0013	−19.27*
South	−0.1116	−10.69*	−0.1834	−10.55*	−0.2817	−11.77*	−0.2803	−11.00*
Gub. Elec.	0.0043	0.38	0.0063	0.33	0.0003	0.02	0.0019	0.09
Closing date	0.0011	0.30	0.0007	0.11	−0.0012	−0.13	−0.0213	−14.65*
Closing date × educ.	−0.0033	−2.18*	−0.0046	−1.79**	−0.0055	−1.45	—	—
Closing date × educ-sq.	0.0003	1.88**	0.0004	1.32	0.0002	0.59	—	—
α	—	—	—	—	0.4150	12.81*	0.4242	13.30*
$P^*(\alpha)$					.40		0.40	
Number of cases	98,857		98,857		98,857		98,857	
Percent voting	67.01		67.01		67.01		67.01	
Correctly predicted	71.04		71.08		71.30		71.30	
Log-likelihood	−55,372		−55,331[c]		−55,283		−55,289	

Source: Data are from the 1984 Current Population Survey.

$^*p < .05$, two-tailed test; $^{**}p < .10$, two-tailed test.

[a] Model specification from Nagler (1991), modified from Wolfinger and Rosenstone (1980), including interactive terms.

[b] Model specification from Wolfinger and Rosenstone (1980).

[c] Log-likelihood from the restricted logit specification is −55,335.

Table 6. Scobit and Logit Correct Predictions for Individuals with Different Levels of Education

Years of Education	Observed Voting Rate	Unrestricted Specification[a] Percent Correctly Predicted Logit	Scobit	Scobit Minus Logit	Restricted Specification[b] Percent Correctly Predicted Logit	Scobit	Scobit Minus Logit
0–4	39.49	60.51	60.38	−0.13	60.44	60.44	0.00
5–7	51.72	57.76	59.40	1.64	57.93	59.50	1.63
8	59.01	64.20	64.33	0.13	64.46	64.33	−0.13
9–11	50.31	65.44	65.65	0.21	65.44	65.66	0.22
12	63.60	68.30	68.28	−0.02	68.28	68.30	0.02
1–3 college	72.85	71.20	71.89	0.69	71.18	71.90	0.72
4 college	83.35	83.35	83.35	0.00	83.35	83.35	0.00
5+ college	88.59	88.59	88.59	0.00	88.59	88.59	0.00

[a]Values from the unrestricted specification—includes multiplicative interactive terms.

[b]Values from the restricted specification—no interactive terms.

reject the hypothesis that $\alpha = 1$ at the 99% level of confidence.[7] Using the formula in equation (11), the α value of .42 translates into a value of P^* of .40. Using our estimate for the standard error of α, we can conclude with 95% confidence that $\alpha \in (.35, .48)$, and that $P^* \in (.38, .42)$. Thus the logit and probit assumption that those voters with initial probability of .5 of voting are most sensitive to change is incorrect. Hence, the assumption implicit in adopting logit or probit—that those individuals with initial probability of voting of .5 are more sensitive to stimulus—is wrong. *It is individuals less likely to vote who are more sensitive to stimulus.*

This suggests testing the fit of the two models, scobit and logit, at different levels of P_i. Since based on the estimate of α and P^* the logit model appears to be misspecified, we would expect to see a pattern of better prediction by scobit at some levels of P. Table 6 reports the

[7]Scobit estimates on the same model without interactive terms also yielded an estimate of α of .42. Estimating a different specification including family income and having no squared or interactive terms yielded an estimate for α of .59 ($P^* = .44$); that the values are so close is encouraging, suggesting that the scobit estimator is robust with respect to specification errors.

percentage of correct predictions by logit and scobit at different levels of education—education serves as a proxy here for changing P—for both the unrestricted (multiplicative interactive terms included) and the restricted (no explicit interactive terms included) models. Recall that Figure 3 indicated that logit would differ from scobit most in the middle and near the tail of the distribution. The areas in which scobit outperforms logit are as Figure 3 would suggest. For the group with probability of voting close to .5 (education level, 5–7), the prediction gap, the difference between the percentage of correct predictions by scobit and logit, is 1.64 and 1.63 in the respective sets of models. The gap decreases as the probability of members of each group voting drops and as the probability of members of each group voting rises, but the gap then increases again for those individuals with probability of voting near .73 (education level, 1–3 years of college). The probabilities converge again at the upper tails. The relatively small difference in correct prediction rates is also suggested by Figure 3.A: with $\alpha = .5$, the two curves are fairly close. Since $\hat{\alpha}$ here is .42, we would not expect the curves to diverge very much.

The question now is how the scobit and logit models differ in predictions of effect of changes in independent variables. The first two columns of Table 7 report the effect estimated by the unrestricted logit and scobit

Table 7. Scobit and Logit Estimates for the Effect of Eliminating the Closing Requirement for Individuals with Different Levels of Education

	Unrestricted Specification[a]		Restricted Specification[b]	
Years of Education	Logit	Scobit	Logit	Scobit
0–4	.022	.028	.081	.094
5–7	.044	.046	.081	.087
8	.053	.050	.069	.069
9–11	.066	.065	.070	.072
12	.067	.067	.064	.064
1–3 college	.064	.063	.058	.055
4 college	.044	.041	.040	.033
5+ college	.023	.024	.022	.018

Note: Probability of voting is calculated for each individual, then closing is set to zero, and a second, hypothetical probability is calculated. Table entries are the mean differences between these two numbers for each education category.

[a]Values from the unrestricted specification—includes multiplicative interactive terms.

[b]Values from the restricted specification—no interactive terms.

models of eliminating the closing requirement. A predicted probability of voting was calculated for each individual. Then, a second hypothetical probability was calculated with the value of *closing* set to zero. The difference between these two numbers was then averaged over all individuals in each education group and reflects the effect of eliminating the closing requirement (Wolfinger and Rosenstone 1980). The scobit model predicts a .028 change in the probability of voting for the group with the lowest level of education (0–4 years), while logit predicts a change of only .022. The difference in the two predictions is caused by the differing shapes of the scobit and logit response curves. Since the group with the lowest level of education has a voting rate of 39.34%, and $P^* = .40$, we would expect scobit to produce larger estimates than logit for this group: this is the group for whom the scobit curve is steepest. I now turn to the interactive effects of education and closing.

Estimates of Variable-Specific Interactive Effects

The shape of the response curve as revealed by the estimates for α shows that, ceteris paribus, poorly educated individuals will be more sensitive to stimuli, since they will be closer to P^*. However, beyond the interaction explicit in the functional form between *all* variables, we might have theoretical reason to postulate a "variable-specific" interaction between some combination of independent variables. In the case of voter turnout, Wolfinger and Rosenstone (1980) argued that voting laws would be particularly difficult for poorly educated individuals to deal with. If so, we would expect estimates of $\Delta P/\Delta closing$ for poorly educated persons to be augmented when including such variable-specific interaction in the model. The variable-specific interaction should add to the "heightened sensitivity" to stimulus of poorly educated people that we are able to measure.

The basic idea behind interactive effects is straightforward.[8] Consider the variables x_k and x_j. We say that there are interactive effects between x_j and x_k if $\partial P_i/\partial x_k$ is a function of x_j. Now obviously in nonlinear models such as those being considered, $\partial P_i/\partial x_k$ is *always* a function of x_j, since even in the simplest specification $\partial P_i/\partial x_k = f(-X_i\beta)\beta_k$, and $X_i\beta$ includes x_j. Generally, if $\partial P/\partial x_k = \Gamma(x_k, P_i)$, where Γ is any function, and $\rho(x_j, P_i) \neq 0$, then we have interaction imposed by functional form between x_k and x_j. However, specifying the interactive terms in the underlying linear form allows for a more precise description of the interaction. If adding the variable $x_{jk} = g(x_k, x_j)$ to the set of variables $x_1, \ldots,$

[8]See Jaccard, Turrisi, and Wan (1990) for further discussion of interaction in linear models.

x_K leads to an improvement in the model, then I would argue that we have "variable-specific" interaction between x_j and x_k, as opposed to interaction imposed between all the variables by the functional form of the model. Including the interactive term $g(x_k, x_j)$ in the underlying linear form allows for $\partial P_i/\partial x_k$ to depend upon values of x_j beyond their impact upon the sum $X_i\beta$. Consider the case where $x_{jk} = g(x_k, x_j) = x_j \times x_k$ is included in the underlying model. Then, equation (5) becomes:

$$\frac{\partial P_i}{\partial x_k} = f(-X_i\beta)(\beta_k + \beta_{jk}x_{ji}). \quad (9)$$

To evaluate this, it is necessary to evaluate both the linear and nonlinear terms.

A simple test for the existence of, though not the direction of, "variable-specific" interactive effects is a log-likelihood ratio test comparing the unrestricted and restricted models, where the restricted model omits $g(x_k, x_j)$. In the present case, two variables are omitted: *closing* × *education* and *closing* × *education*^{2}. Results of the restricted scobit model are presented in columns 7 and 8 of Table 5. According to the log-likelihood statistics, we can reject the restricted scobit model versus the unrestricted scobit model with 99% confidence. However, the difference in values of the log-likelihoods (55,282 versus 55,289) is small given the number of observations, and the percentage of correct predictions by the two models is identical (71.30). These things do not suggest a very meaningful substantive difference between the two models. Looking at the coefficients of the specific interactive terms in the model, we can see that they do not individually reach traditional levels of significance ($t = -1.45$ and $t = 0.59$).

Comparing the estimates of the unrestricted and restricted scobit models tells us what effect the variable-specific interaction has. The restricted model predicts the largest effect ($\Delta P = .094$) for the least-educated group, and the smallest effect ($\Delta P = .018$) for the most-educated group. This is a ratio of over 5:1 for the magnitude of the effects. Alternatively, once the multiplicative interactive terms are added, the unrestricted scobit model predicts the largest change for high school graduates ($\Delta P = .067$), with smaller changes at both extremes ($\Delta P = .028$ and $\Delta P = .024$). Thus, the inclusion of variable-specific interaction into the model *attenuates* the tendency of poorly educated individuals to be sensitive to stimuli.

This attenuation effect can also be seen by examining the effect of changes in closing on the underlying variable Y^*. Using the coefficient estimates in Table 5, we have:

$$\frac{\partial^2 Y^*}{\partial closing\, \partial education} = -.0055 + .0004 \times education. \tag{10}$$

Since the variable *education* is bounded between one and eight, this is always negative. Hence, as education increases, the effect of increasing the closing date on Y^* decreases (i.e., becomes more negative); hence, changes in closing have *more* of an effect on Y^* for more highly educated persons. This analysis based on the scobit coefficients is consistent with Nagler's analysis of probit estimates (1991).

Thus, we see that attempts to find variable-specific interactive effects between education and closing remain futile. The interactive variables are not individually significant; their effect on the underlying variable Y^* is perversely signed; and a restricted model without variable-specific interaction compared to an unrestricted model with variable-specific interaction between education and closing produces estimates of effects indicating that the natural tendency of poorly educated individuals to be highly sensitive to stimulus is *attenuated* by the variable-specific interaction. Individuals with lower levels of education are more sensitive to changes in the closing date than individuals with higher levels of education. However, as we have seen, this is because of the greater sensitivity of poorly educated individuals to *any* stimuli; it is *not* because of a peculiar link between education and ability to register early.

Conclusion

I began by identifying a limitation of probit and logit estimators. A technical assumption implicit in the models, that the distribution of the disturbances is symmetric about $F(z) = .5$, has a serious substantive implication: that individuals with $P_i = .5$ are most sensitive to changes in independent variables. I attempted to explain why we have important reasons to want to test this substantive assumption. The scobit, or more descriptively "skewed-logit," estimator developed here overcomes the limitations of probit and logit and offers a means to test this assumption. Monte Carlo results have shown that the scobit estimator can be used even when logit is applicable. And estimates based on actual data have shown that the substantive assumption in question is violated in the case of voter turnout.

It would, however, be a mistake to generalize this result to other substantive cases. The fact that respondents with an initial reported probability of voting of .40 are most sensitive to changes in stimuli does not suggest that individuals with a high or low initial probability of voting Democratic are those most sensitive to stimuli. That the one example considered here generated a P^* close to .4 does not suggest that $\alpha = .42$

represents the true model for every other binary-choice phenomenon in the real world. It remains to be seen whether this would hold in other cases such as candidate choice. The true model in other cases may look more like logit, or may suggest that logit is even more inappropriate. Each substantive case would represent a different empirical question, all of which are beyond the scope of this paper.

As Wolfinger and Rosenstone (1980) initially argued, individuals with less education *will* be more affected by changes in registration laws. However, this is *not* because of a particular link between education and registering early; it is because individuals with less education are more sensitive to any changes likely to increase turnout. In fact, Wolfinger and Rosenstone were even "more" correct than they realized, since the scobit estimates indicated that logit and probit will underestimate the sensitivity to stimulus of persons with extremely low initial probabilities of voting.

Thus, scobit is proved to be a useful estimator when we are explicitly concerned with heterogeneity of respondents and interaction effects. It is useful when, *for substantive reasons,* we are unwilling to make the assumptions of the logit and probit models. Given the closeness of logit and scobit predictions, a researcher not especially interested in questions of interaction between variables or heterogeneity of respondents would probably not feel compelled to compute scobit estimates. However, I argue that scobit is an essential estimator if we want to test for heterogeneity of respondents and interactive effects when dealing with dichotomous dependent variables. The purpose of empirical analysis is to test hypotheses, and if the proof of our hypothesis is imposed upon our results by our statistical model then no test is being conducted.

Manuscript submitted 23 November 1992
Final manuscript received 25 May 1993

APPENDIX A
Proof of Proposition 1

Define $\dot{P}_i$ as $\partial P_i/\partial z$, and define P^* as the value of P that maximizes $\dot{P}_i$. The goal here is to solve for P^* in terms of α, and then show that choosing an appropriate value of α will enable P^* to take on any value within the range (0, 1). Substituting the Burr-10 distribution into equation (5) shows that P^* maximizes:

$$\dot{P}_i = \frac{\partial[1 - (1 + e^z)^{-\alpha}]}{\partial z}. \tag{11}$$

Or P^* maximizes:

$$\dot{P}_i = \alpha(e^z)(1 + e^z)^{-(\alpha+1)}. \tag{12}$$

We can solve for the value z^* where $\dot{P}$ is maximized by differentiating equation (8) with respect to z and setting the result equal to zero. Since $f(-z)$ is concave downward $\forall$ z, satisfying the first-order condition will guarantee that we have a maximum. This gives the first-order condition:

$$\frac{\partial \dot{P}_i}{\partial z} = \alpha(e^z)(1 + e^z)^{-(\alpha+2)}(1 - \alpha e^z) \tag{13}$$
$$= 0.$$

A maximum is found where:

$$1 - \alpha e^z = 0,$$

or

$$z^* = \log\left(\frac{1}{\alpha}\right) \tag{14}$$

This is the value of z where $(\dot{P}_i|\alpha)$ is maximized $\forall$ α. Substituting this into equation (4), we can express P^* in terms of α as:

$$P^* = 1 - \left(1 + \frac{1}{\alpha}\right)^{-\alpha} \tag{15}$$

Hence $P^* \rightarrow 0$ as $\alpha \rightarrow 0$; but in the limit we can see that $P^* \rightarrow (1 - (1/e))$, or approximately .63, as $\alpha \rightarrow \infty$. Hence, $\forall$ $k \in (0, .63)$ $\exists$ α s.t. $P^* = k$. However, $\forall$ $k > .63$, there does *not* exist α s.t. $P^* = k$. While this distribution would not satisfy the criteria of being able to generate any $P^* \in (0, 1)$, a recoding technique is available that would allow the distribution to be viable for $P^* > .63$.

Consider the case where $P^* > .63$. If $P^* = \text{prob}(Y_i = 1) > .63$, then $\text{prob}(Y_i = 0) = (1 - P^*) < .37$. This suggests recoding the dependent variable to reverse the 0/1 coding and recalculating the parameter estimates. Define $\tilde{Y}_i$ as follows: $\tilde{Y}_i = 0$ if $Y_i = 1$; $\tilde{Y}_i = 1$ if $Y_i = 0$. Now define $\tilde{P}_i = \text{prob}(\tilde{Y}_i = 1) = \text{prob}(Y_i = 0) = 1 - \text{prob}(Y_i = 1) = 1 - P_i$. Thus, the value of $\tilde{P}$ where $\partial\tilde{P}/\partial z$ is maximized ($\tilde{P}^*$) will just be $1 - P_i^*$. Since $P_i^* > .67$, it follows that $1 - P_i^* < .37$, thus $\tilde{P}^* < .37$. Hence, with the data coded in this manner $\exists$ α s.t. $\tilde{P}^* \in (0, 1)$; and the Burr-10 distribution could be used when $P^* \in (1/e, 1)$. QED

Since the coding of data as zero or one is an arbitrary decision regarding our classification of events, the "recoding" involves no loss of generality. Following the coding change, we would simply have to be precise as to the "new" meaning of our dependent variable. Thus, determining the correct distribution for $P^* \in (0, 1)$ would require running two sets of maximum likelihood calculations. However, aside from computational and coding complexity, there is no loss of efficiency.

APPENDIX B
Likelihood Function for the Burr-10 Distribution

Following the derivation of the general model for dichotomous dependent variables, the likelihood function is defined as:

$$L = \Pi F(-X_i\beta)^{(1-y_i)}(1 - F(-X_i\beta))^{y_i}. \tag{16}$$

The log of the likelihood function is then given by:

$$\text{Log}\, L = \Sigma(1 - y_i)\log[F(-X_i\beta)] + \Sigma y_i \log[1 - F(-X_i\beta)], \tag{17}$$

where F represents the Burr-10 distribution, or:

$$F(-X_i\beta) = (1 + e^{(X_i\beta)})^{-\alpha}. \tag{18}$$

The gradient is given by the vector:

$$\frac{\partial \log L}{\partial \beta_k} = ((y_i/(1 - F(-X_i\beta;\alpha))) - ((1 - y_i)/F(-X_i\beta;\alpha))) \times f_\beta(-X_i\beta;\alpha) \times x_k, \tag{19}$$

and

$$\frac{\partial \log L}{\partial \alpha} = ((-y_i/(1 - F(-X_i\beta;\alpha))) + ((1 - y_i)/F(-X_i\beta;\alpha))) \times f_\alpha(-X_i\beta;\alpha), \tag{20}$$

where

$$\begin{aligned} f_\beta(-X_i\beta;\alpha) &= \frac{\partial F(-X_i\beta;\alpha)}{\partial\beta} \\ &= -\alpha \times e^{X_i\beta} \times (1 + e^{X_i\beta})^{-(\alpha+1)} x_{ki} \end{aligned} \tag{21}$$

and

$$\begin{aligned} f_\alpha(-X_i\beta;\alpha) &= \frac{\partial F(-X_i\beta;\alpha)}{\partial\alpha} \\ &= -\log[1 + e^{(X_i\beta)}] \times F(-X_i\beta;\alpha) \end{aligned} \tag{22}$$

APPENDIX C
SHAZAM Code for Scobit Estimation

The following SHAZAM code can be used to generate scobit estimates for a model with two independent variables ($x1$ and $x2$) and a binary dependent variable y:

```
nl 1 / ncoef=4 logden
eq (1-y)*log((1+exp(b0+b1*x1+b2*x2))**(-1*alpha)) &
+y*log(1-(1+exp(b0+b1*x1+b2*x2))**(-1*alpha))
coef b0 1 b1 1 b2 1 alpha 1
end
```

Note: The symbol "**" denotes exponentiation in SHAZAM; the symbol "&" denotes that the equation continues to the next line.

REFERENCES

Aldrich, John H., and Forrest D. Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Beverly Hills: Sage.

Burr, Irving W. 1942. "Cumulative Frequency Functions." *Annals of Mathematical Statistics* 13:215–32.

Härdle, Wolfgang. 1990. *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Harvey, Andrew C. 1981. *The Econometric Analysis of Time Series*. Oxford: Philip Allan.

Jaccard, James, Robert Turrisi, and Choi K. Wan. 1990. *Interaction Effects in Multiple Regression*. Newbury Park, CA: Sage.

Jacobson, Gary C. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34:334–62.

King, Gary. 1989. *Unifying Political Methodology*. Cambridge: Cambridge University Press.

Maddala, G. S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Nagler, Jonathan. 1991. "The Effects of Registration Laws and Education on Voter Turnout." *American Political Science Review* 85:1393–1405.

Prentice, Ross L. 1976. "A Generalization of the Probit and Logit Methods for Dose Response Curves." *Biometrics* 32:761–68.

White, Kenneth J., S. Donna Wong, Diane Whistler, and Shirley A. Huan. 1990. *SHAZAM User's Reference Manual Version 6.2*. New York: McGraw-Hill.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven: Yale University Press.