Political Analysis, 8:2

Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models

Lee Sigelman and Langche Zeng
The George Washington University

Political scientists are making increasing use of the Tobit and Heckit models. This paper addresses some common problems in the application and interpretation of these models. Through numerical experiments and reanalysis of data from a study by Romer and Snyder (1994), we illustrate the consequences of using the standard Tobit model, which assumes a censoring point at zero, when the zeros are not due to censoring mechanisms or when actual censoring is not at zero. In the latter case, we also show that Greene's (1981) well-known results on the direction and size of the bias of the OLS estimator in the standard Tobit model do not necessarily hold. Because the Heckit model is often used as an alternative to Tobit, we examine its assumptions and discuss the proper interpretation of the Heckit/Tobit estimation results using Grier and co-workers' (1994) Heckit model of campaign contribution data. Sensitivity analyses of the Heckit estimation results suggest some conclusions rather different from those reached by Grier et al.

1 Introduction

INCREASING USE IS being made of the Tobit and Heckit regression models to analyze a wide range of political phenomena, including campaign contributions (Chappell 1982; Grier and Munger 1993; Grier et al. 1994; McCarty and Rothenberg 1996), PAC activities (Romer and Snyder 1994), vote choice (Herron 1998), the president's use of military force (Meernik 1994; Morgan and Bickers 1992; Wang 1996), the occurrence of political protest (Roncek 1992; Walton and Ragin 1990), and the determinants of police violence (Jacobs and O'Brien 1998).

Confusion reigns, however, about the proper use of these models and the appropriate interpretation of findings from them. Theoretically the standard Tobit model is applicable only if the underlying dependent variable contains negative values that have been censored to zero in the empirical realization of the variable. In practice, though, the Tobit model is routinely employed when the values of the observed dependent variable are exclusively nonnegative and are clustered at zero, irrespective of whether any censoring has occurred. The Heckit model has emerged as the de facto default alternative to Tobit when values

Authors' note: The authors thank Chris Achen, Richard Timpone, and the anonymous reviewers for their helpful comments and suggestions and Kevin Grier, Michael Munger, Thomas Romer, and James Snyder for providing the data on which some of the analyses reported are based. Langche Zeng thanks the National Science Foundation for research support (Grant SBR-9753126).

cluster at zero due to selection bias rather than censoring, but applications of the Heckit model have proven problematic as well.

In this paper, we seek to heighten awareness of the conditions under which these models should and should not be used and of the consequences of using them incorrectly, and to contribute to the appropriate use of these models by illustrating the proper interpretation of results. We focus in Section 2 on the Tobit model and in Section 3 on the Heckit model. Conclusions are contained in Section 4.

2 The Tobit Model

In a probit model the variable of theoretical interest, y^*, is unobserved; what is observed is a dummy variable, y, which takes on a value of 1 if y_i^* is greater than 0, and 0 otherwise. In contrast, Tobin (1958) devised what became known as the Tobit (Tobin's probit) or censored normal regression model for situations in which y is observed for values greater than 0 but is not observed (that is, is censored) for values of zero or less.

The standard Tobit model is defined as

$$\begin{aligned} y_i^* &= \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i \\ y_i &= y_i^* \quad \text{if} \quad y_i^* > 0 \\ y_i &= 0 \quad \text{if} \quad y_i^* \le 0 \end{aligned} \tag{1}$$

where y_i^* is the latent dependent variable, y_i is the observed dependent variable, $\mathbf{x}_i$ is the vector of the independent variables, $\boldsymbol{\beta}$ is the vector of coefficients, and the ϵ_i's are assumed to be independently normally distributed: $\epsilon_i \sim N(0, \sigma)$ (and therefore $y_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \sigma)$).[1] Note that observed 0's on the dependent variable can mean either a "true" 0 or censored data. At least some of the observations must be censored data, or y_i would always equal y_i^* and the true model would be linear regression, not Tobit.

For this data configuration, it is well known that OLS estimators are biased downward (e.g., Greene 1997). A simple Monte Carlo experiment illustrates the bias. We generate the Tobit y using the latent index function $y^* = \alpha + x + e$, with both x and e being standard normal and with the value of α being set to 1, 0, and -1 so that the proportion of data uncensored is 76%, 50%, and 24%, respectively.[2] Each data set has 1000 observations, and the experiment is replicated 100 times. Table 1 shows the mean OLS estimates of the coefficient for x (standard deviations in parentheses), the true value of which is 1. With 24% of the observations uncensored, the mean of the estimated $\boldsymbol{\beta}$ values is .24; with 50%, .51; and with 76%, .77. The symmetry of these results is, of course, hardly coincidental, for it has been proved that the ratio of the OLS estimates to the maximum likelihood estimates (which are consistent and in this case close to 1) approximates the proportion of the data uncensored (Goldberger 1972), a result made well known by Greene (1981).

Maximum-likelihood estimation of the Tobit model is straightforward. Let $f(.)$ and $F(.)$ denote the density function and the cumulative density function for y^*. Then the model implies that the probabilities of observing a non-zero y and a zero y are $f(y)$ and

[1] Hereafter we omit subscripts where harmless.

[2] This is so because $x + e \sim N(0, \sqrt{2})$. Thus, for example, for $\alpha = 1$, $p(y^* > 0) = p(1 + x + e > 0) = p(x + e > -1) = .76$.

P1: FBI/FKP P2: FBW
WV002-05 December 16, 1999 11:36

Table 1 Bias of OLS estimates on censored data

Percentage of data uncensored	*Mean of OLS estimates of the x coefficient (SD)*	*True value of the x coefficient*
24	.24 (.020)	1.0
50	.51 (.025)	1.0
76	.77 (.027)	1.0

$p(y^* < 0) = F(0)$, respectively. The log-likelihood function for the model is therefore

$$\ln L = \ln\left(\prod_{y_i>0} f(y_i) \prod_{y_i=0} F(0)\right) = \sum_{y_i>0} \ln f(y_i) + \sum_{y_i=0} \ln F(0) \quad (2)$$

Because y^* is normally distributed (as the $\epsilon' s$ are normally distributed), $f(.)$ and $F(.)$, and therefore the log-likelihood function, can be reexpressed in terms of the density function and the cumulative density function of the standard normal distribution, $\phi(.)$ and $\Phi(.)$, and the log-likelihood function can be written in the familiar form:

$$\ln L = \sum_{y_i>0}\left(-\ln\sigma + \ln\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\right) + \sum_{y_i=0} \ln\left(1 - \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\right) \quad (3)$$

Maximum likelihood estimation can then proceed in the usual fashion.[3]

To interpret the estimation results, the marginal effects of the independent variables on some conditional mean functions should be examined. In the familiar OLS model $y = \mathbf{x}\boldsymbol{\beta} + \epsilon$, there is only one conditional mean function, $E(y) = \mathbf{x}\boldsymbol{\beta}$, and $\partial E(y)/\partial \mathbf{x}_k = \boldsymbol{\beta}_k$, where $\mathbf{x}_k$ is the kth independent variable. This makes interpretation easy: $\boldsymbol{\beta}_k$ measures the marginal effect on y of the kth independent variable. In the Tobit model, though, there are three different conditional means: those of the latent variable y^*, the observed dependent variable y, and the uncensored observed dependent variable $y \mid y > 0$. Accordingly, interpretation depends on whether one is concerned with the marginal effect of $\mathbf{x}$ on y^*, y, or $y \mid y > 0$. For example, in analyzing data on campaign contributions, one would be interested in y^* if the goal were to understand the underlying *propensity* to donate, y to understand the determinants of the *actual* amount of donations by contributors and noncontributors alike, and $y \mid y > 0$ to understand the amount of donations by contributors alone. Once one determines which marginal effect one is interested in, one simply examines the marginal effects of $\mathbf{x}$ on the appropriate conditional expectations. The three marginal effect expressions are derived using standard results on moments of truncated/censored normal distributions (see, e.g., Greene 1997, pp. 962–963), as follows:

$$\frac{\partial E(y^* \mid \mathbf{x})}{\partial \mathbf{x}} = \boldsymbol{\beta} \quad (4)$$

[3]The model is described in most econometrics texts, e.g., Greene (1997). Our purpose here is not to provide a full exposition of the Tobit model, but only to highlight its most essential aspects.

$$\frac{\partial E(y \mid \mathbf{x})}{\partial \mathbf{x}} = \boldsymbol{\beta}\Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right) \tag{5}$$

$$\frac{\partial E(y \mid y > 0, x)}{\partial \mathbf{x}} = \boldsymbol{\beta}\left(1 - \delta\left(-\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\right) \tag{6}$$

where $\delta(\alpha) = \lambda(\alpha)(\lambda(\alpha) - \alpha)$, $\lambda(\alpha) = \phi(\alpha)/(1 - \Phi(\alpha))$, and $\alpha = -(\mathbf{x}\boldsymbol{\beta}/\sigma)$.[4] Clearly, only for the latent index y^* can $\boldsymbol{\beta}$ be interpreted as the marginal effects of the independent variables.[5]

To reiterate, the standard Tobit model assumes, among other things, that the dependent variable is censored at zero. *If no censoring has occurred or if censoring has occurred but not at zero, then the standard Tobit specification is inappropriate.* Maddala (1992, p. 341), for example, sounds a clear warning against using the Tobit model when no censoring has occurred:

> Every time we have some zero observations in the sample, it is tempting to use the Tobit model. However, it is important to understand what the model . . . really says. What we have . . . is a situation where y_i^* can, *in principle,* take on negative values. However, we do not observe them because of censoring. Thus the zero values are due to nonobservability. This is *not* the case with automobile expenditures, hours worked, or wages. These variables cannot, in principle, assume negative values. The observed zero values are due not to censoring, but due to the decisions of individuals. In this case the appropriate procedure would be to model the decisions that produce the zero observations rather than use the Tobit model mechanically.

One might, for example, consider using a Tobit model to study congressional campaign contributions. Because many potential donors contribute exactly $0 to a particular candidate, there is likely to be a large cluster of observations at $0, along with a range of observations with positive values. Is this a data configuration for which a Tobit model is appropriate? The answer depends on one's assumptions about the nature of the decision-making problem facing PACs. Does a PAC decide on how much it *prefers* or *wishes* to contribute, or does it decide on whether or not to contribute and then, if the first decision is affirmative, decide on the exact amount of its contribution? If one is willing to *assume* the former, then this data configuration can be modeled via Tobit, for a PAC might wish that it could make a "negative contribution" to a disliked candidate by taking dollars away. In that case, the underlying propensity to contribute to a particular candidate can be imagined to include negative as well as positive values, with $0 representing a censored negative observation. However, if one believes that in practice it is not likely that a PAC would be contemplating any impossible preferences, but rather would be focusing on a "give or not" decision first, then clearly the Tobit model is inappropriate, because no censoring would be involved. A $0 amount would be the result of binary decision making rather than censoring.

Unfortunately, this elementary point is routinely ignored. Consider, for example, the following recent applications of the Tobit model.

- In a study of PAC contributions to congressional candidates, Romer and Snyder (1994) counted, for each candidate, the number of PACs that changed their support from one

[4] Equation (5) can be decomposed into two parts for ease of interpretation (McDonald and Moffitt 1980). Roncek (1992) provides an example.

[5] There can be cases in which the mean of the latent y^* is of central interest, but when the data are censored the mean of the observed y is usually of greater interest. In analyzing campaign contribution data, for example, interest usually centers on the determinants of the observed contributions, not of the unobserved potential contributions.

election to the next—either by contributing to a candidate in the first campaign of a two-campaign cycle but not in the second (“dropping”), by contributing in the second campaign after not contributing in the first (“adding”), or by either dropping or adding (“changing”). Romer and Snyder (1994, p. 756) employed the Tobit procedure “because these variables are bounded below by zero, and several of them frequently take on the limit value of zero.” Although these “look like” Tobit data, in that each y consists of a cluster of zero values and a set of positive values, no censoring has occurred. It follows that these are not appropriate data for a Tobit model. They are typical event count data, for which models like Poisson regression (King 1989) are preferable.

- To test the “diversionary” hypothesis that “The probability of a state engaging in aggressive foreign policy behavior increases when the leadership of a government is faced with decreased support among members of its ruling coalition,” Morgan and Bickers (1992) probed the impact of the president’s popularity among members of his own party on the use of force in U.S. foreign policy. More specifically, Morgan and Bickers measured the number of days from a given reading of presidential popularity until the next incident in which the United States threatened, displayed, or actually used military force against another country, 1953–1976. Because the lower bound of y, the elapsed time between a survey and the next incident of military force, was zero, they modeled this relationship via Tobit. Here again, though, the boundedness of y was not a result of censoring. Rather, it reflected the very nature of y, which logically could not assume a value of less than zero for the simple reason that a duration cannot be negative. More appropriate procedures for analyzing duration data are readily available (Box-Steffensmeier and Jones 1997).
- Wang (1996) also analyzed the president’s use of force in the conduct of American foreign policy. But whereas Morgan and Bickers focused on how long it took the president to respond to his standing in the polls, Wang focused on the severity of the president’s response. He coded each presidential response on a scale that ranged from 0, denoting compliance with an opponent’s demands, to 1, signifying violent military action, with nine intermediate responses unequally spaced between 0 and 1. “Because the dependent variable is limited between values of 0 and 1,” Wang (1996, p. 78) supplemented his main statistical analyses with a Tobit analysis of these data. Again, though, no censoring was involved; 0 was a naturally occurring and relatively frequent value in the data, not the value at which certain naturally occurring values had been censored. Accordingly, a more appropriate model would be one that restricted the range of y without assuming censoring, such as a nonlinear regression like $y = 1/(1+e^{-\mathbf{x}\boldsymbol{\beta}})$, which is the logit form and always takes on values between 0 and 1.

These are just a few of the numerous instances we have observed in analyses of political data where the Tobit model has been used in the absence of censoring—a situation for which it was not intended and is not appropriate. Before we proceed any further, it is important to establish more specifically what the problem is when one employs Tobit in such situations and what the consequences of such misapplication are.

The core of the problem is simple. If there has been no censoring, the likelihood function is changed, and Eq. (2) or Eq. (3) no longer applies. For example, if y is an event count variable generated from a Poisson process, then the probability of $y = y_i$ for all y_i, positive or zero, is given by the Poisson model, and the likelihood function for the data should be

$$\ln L = \ln \prod_{i=1}^{N} \frac{e^{-e^{\mathbf{x}_i \boldsymbol{\beta}}}(e^{\mathbf{x}_i \boldsymbol{\beta}})^{y_i}}{y_i\,!} \tag{7}$$

where we assume that $E(y_i) = e^{\mathbf{x}_i \boldsymbol{\beta}}$. Obviously, the $\boldsymbol{\beta}$ estimated by maximizing this likelihood function will be very different from the $\boldsymbol{\beta}$ estimated by maximizing the Tobit log likelihood function in Eq. (3).

To illustrate the consequences of using the Tobit model when no censoring has occurred, consider the count data that Romer and Snyder (1994) analyzed using Tobit.[6] These data feature a cluster of $0 values and positive integers, where the zeros result from the underlying data-generation process rather than censoring. For such data Poisson regression models are more appropriate than Tobit models.

Table 2 compares the estimation results from Poisson models (Table 2-2) with those for the Tobit models (Table 2-1) that Romer and Snyder reported in their Table 3. Table 2 shows the results for six specifications of each model, which differed from one another primarily in terms of the definition of the dependent variable: the number of PACs that "added," "dropped," or "changed" donations to a candidate, as explained earlier, and then the same three variables, but considering only donations of at least $500 rather than all donations. The underlying model was

$$\begin{aligned} y = c &+ \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 \\ &+ \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} \end{aligned}$$

where y is the number of adds, drops, or changes for a given candidate in a year, x_1 (W & M) is a dummy variable indicating whether a member of Congress moved to the Ways and Means Committee in the interim, x_2 (Apprul), x_3 (E & C), and x_4 (Other2) are parallel dummies for Appropriations or Rules, Energy and Commerce, and two different nonexclusive committees, respectively, x_5 (Leader) is a dummy variable for moving into a committee leadership post, x_6 (DVS) is the (signed or absolute) change in the incumbent member's share of the two-party vote,[7] x_7 (DCREC) is the (signed and absolute) change in total donations to the challenger, x_8 (NPAC) is the number of PACs contributing to the member, and x_9 (Y82-84), x_{10} (Y84-86), and x_{11} (Y86-88) are dummy variables identifying the election years in question (1982–1984, 1984–1986, or 1986–1988).[8]

To facilitate comparison, we follow Romer and Snyder's practice of marking variables that are significant at the .01 level with two asterisks, or at the .05 level with one asterisk. Inspection of Table 2 reveals that more significant effects surfaced in the Poisson models than in the (inappropriate) Tobit models. The most striking differences were for the third and sixth models, which focused on the number of PACs that dropped their contributions from one election to the next. In these Poisson models *every* explanatory variable had a highly significant effect, while in these Tobit models the effects of three or four predictors were nonsignificant. For example, both Tobit models of drops (and five of the six Tobit models overall) failed to uncover the significant effect of the 1982–1984 dummy variable. This variable and its 1984–1986 and 1986–1988 counterparts were in the model "to control for changes in the number of PACs over time and changes in nominal and real PAC budgets, as well as for year-specific factors affecting PAC contributions" that might not be captured by other predictors—most obviously, for the 1982–1984 election pair, the redrawing of

[6]Our focus on the Romer and Snyder data set and below on the Grier et al. data set is not motivated by a conviction that these analyses are unusually problematic; if anything the opposite is the case. Rather, Romer and Snyder and Grier et al. graciously made their data available to us. Their reward is to be singled out for criticism.

[7]For the change dependent variable, the absolute value of x_6 is used, and for the adds and drops, the signed version is used. The same is true for x_7.

[8]For detailed descriptions of these variables, see Romer and Snyder (1994).

Table 2 Tobit model vs. Poisson regression on Romer–Snyder's (1994) count data

Independent variable	*Dependent variable*					
	Changes	*Adds*	*Drops*	*Changes(500)*	*Adds(500)*	*Drops(500)*
2-1: Tobit results (replication of Romer and Snyder's Table 3)						
W & M	29.14**	45.91**	−17.43**	9.88	26.31**	−17.69**
	(3.60)	(4.52)	(3.94)	(5.31)	(5.02)	(4.17)
Apprul	−.52	−5.00	4.76	−7.70	−10.79*	3.60
	(3.09)	(3.88)	(3.38)	(4.56)	(4.31)	(3.58)
E & C	17.94**	35.17**	−16.87**	−9.81	4.82	−14.16**
	(4.35)	(5.47)	(4.77)	(6.42)	(6.08)	(5.05)
Other2	4.28*	−.10	4.23*	−3.85	−4.97*	1.03
	(1.74)	(2.2)	(1.91)	(2.56)	(2.44)	(2.03)
Leader	.53	8.37**	−7.90**	3.02	12.20**	−10.00**
	(2.54)	(3.19)	(2.79)	(3.75)	(3.55)	(2.97)
DVS(|DVS|)	25.80*	69.15**	−47.45**	38.98*	79.11**	−55.27**
	(10.90)	(11.34)	(9.88)	(16.08)	(12.59)	(10.47)
DCREC(|DCREC|)	.65**	2.61**	−2.07**	1.67**	3.33**	−2.89**
	(.25)	(.29)	(.25)	(.36)	(.32)	(.26)
NPAC	.48**	.28**	.21**	.53**	.34**	.21**
	(.006)	(.007)	(.006)	(.009)	(.008)	(.007)
Y82-84	−2.03	−1.55	−.17	3.89*	1.96	1.84
	(1.28)	(1.61)	(1.41)	(1.89)	(1.80)	(1.49)
Y84-86	−3.08*	−5.1**	2.23	13.19**	4.22*	8.62**
	(1.29)	(1.63)	(1.42)	(1.90)	(1.81)	(1.51)
Y86-88	−7.70**	−11.67**	3.93**	20.02**	4.83**	14.66**
	(1.30)	(1.63)	(1.42)	(1.92)	(1.81)	(1.51)
Const.	6.96**	5.41**	1.64	−24.64**	−12.69**	−12.00**
	(1.24)	(1.52)	(1.32)	(1.84)	(1.69)	(1.41)
2-2: Poisson results on Romer and Snyder data						
W & M	.27**	.58**	−.38**	.16**	.38**	−.49**
	(.02)	(.02)	(.04)	(.02)	(.02)	(.05)
Apprul	.03	−.07	.15**	−.05*	−.19**	.17**
	(.02)	(.03)	(.03)	(.02)	(.03)	(.03)
E & C	.12**	.41**	−.39**	−.15**	.02	−.41**
	(.02)	(.03)	(.05)	(.03)	(.03)	(.05)
Other2	.08**	.03	.14**	−.02	−.07**	.07**
	(.01)	(.02)	(.02)	(.01)	(.02)	(.02)
Leader	.01	.13**	−.17**	.08**	.23**	−.24**
	(.02)	(.02)	(.03)	(.02)	(.02)	(.03)
DVS(|DVS|)	.36**	1.24**	−1.38**	.62**	1.51**	−2.24**
	(.07)	(.08)	(.10)	(.08)	(.08)	(.11)
DCREC(|DCREC|)	.01**	.04**	−.04**	.02**	.05**	−.06**
	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)
NPAC	.005**	.005**	.005**	.005**	.006**	.005**
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Y82-84	−.007	−.04**	.06**	.154**	.096**	.32**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)
Y84-86	−.02	−.10**	.12**	.28**	.15**	.57**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)
Y86-88	−.07**	−.24**	.15**	.33**	.15**	.70**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)
Const.	3.59**	3.06**	2.72**	3.04**	3.68**	1.88**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)

House districts (Romer and Snyder 1994, p. 755). Thus, whereas Romer and Snyder's Tobit analysis led them to conclude that changes during the 1982–1984 interval, including redistricting, did not influence PAC contributions, a more appropriate specification of the model, in the form of a Poisson model, yielded the opposite conclusion. Similarly, the failure of the Tobit version of the third model to turn up a significant effect on drops for the 1984–1986 election pair was again contradicted in the Poisson model.

More substantively, Romer and Snyder also expected PAC contributions to drop when a member moved from one nonexclusive committee to another, not only because these were lateral moves rather than promotions but also because they involved a loss of the seniority and expertise the member had built up on his or her previous committee (Romer and Snyder 1994, p. 749). The nonsignificant effect of this variable in model 6 thus came as a surprise for which Romer and Snyder offered no explanation. According to our reanalysis, had the model been specified as Poisson rather than Tobit, no explanation would have been necessary, for the result would have been as predicted.

Finally, Romer and Snyder also expected that switches to certain committees, but not to others, would affect PAC contributions. In the forefront in this regard were the Ways and Means and Energy or Commerce committees, and as can be seen in Table 2, this expectation was borne out. What Romer and Snyder did not note was a significant effect one way or the other of switching to either of two powerful committees, Appropriations or Rules. In five of the six Tobit models, including both models of drops, such a switch appeared to have no significant effect on PAC contributions. In five of the six Poisson models, though, including both models of drops, such a switch did significantly affect PAC contributions. Interestingly, these significant effects were invariably in the direction of decreasing adds or increasing drops. Thus, a story remains to be told about why joining these committees undermined members' ability to attract PAC contributions—the need for which now arises because the models have been appropriately specified as Poisson rather than Tobit.

Now consider the case in which censoring has occurred, but at a point other than zero (e.g., $y = 0$ if $y^* \leq c$, where $c \neq 0$.) Under these conditions the probability of observing $y = 0$ becomes $P(y^* \leq c) = F(c) \neq F(0)$. Substantively, censoring at a nonzero point is hardly far-fetched, for there is nothing inevitable or magical about zero. For example, in analyses of the dollar value of PAC contributions to candidates (e.g., Grier and Munger 1993; Grier et al. 1994), it makes little sense to assume that a cluster of zero observations reflects censoring at zero. For one thing, as noted above, a PAC contribution can never be negative. For another, it seems highly unlikely that any PAC would donate a miniscule amount, say a dollar or two, to a candidate; that being the case, the true censoring point may be well above \$0, at, say, \$50, \$100, or \$500. Finally, some sort of censoring from above may also occur, as in the form of a limit on the size of the contribution any PAC can make.[9]

In every study we have examined, censoring is assumed to have occurred at zero. If this assumption is incorrect, the likelihood function will differ from what is assumed in the standard Tobit model and estimation via the standard Tobit model will in general be inappropriate. However, depending on how far away c is from zero and how far away zero is from the mean of the data (which together determine the difference between $F(c)$ and $F(0)$), the consequences of using the standard Tobit model may or may not be substantial.

To explore the consequences of using the standard Tobit model when censoring has occurred at a point other than zero, we generated 100 data sets of 1000 observations at

[9]For such data the disequilibrium model seems promising (Maddala 1983).

Table 3 Standard Tobit estimation of data censored at various points (True value of coefficient = 1)

Mean estimate	*SD*	*Data censored at*
1.001	.040	.0
1.162	.045	.5
1.471	.054	1.0
1.954	.086	1.5
2.655	.125	2.0

each of five different censoring points, using the function $y^* = 1 + x + e$, where, as in the simulation of bias in OLS results reported earlier, both x and e were drawn from the standard normal distribution. Table 3 shows the Tobit estimate of the coefficient for x, the true value of which is 1. As expected, with the data censored at zero the Tobit estimate of the coefficient approximates 1; this is exactly as it should be, for in this instance the assumption underlying the standard Tobit model perfectly matches the actual censoring point. However, as the censoring point moves progressively farther from zero, the bias of the Tobit estimates worsens.[10] Censoring at a nonzero point does not in itself pose a threat to correct estimation, for the correct likelihood function can be specified in a straightforward manner, similar to that of the standard Tobit. The problem arises because of the misspecification of the actual censoring value(s), not because of nonzero censoring per se.

We noted earlier that Goldberger (1972; see also Greene 1981) proved that, for data for which the standard Tobit model was intended, the OLS estimators bias downward and the degree of the bias is empirically related to the proportion of data censored. However, Table 4 (which is based on the same simulated data as Table 3) illustrates that when censoring is not at zero, the magnitude of the downward bias of OLS estimates bears no definite relationship to the proportion of data censored. Nor must the bias always be downward. This can be seen in Figs. 1 and 2, which plot two sets of censored data and the OLS regression lines for these data, to be compared with the true regression lines. In Fig. 1, the latent index function is $y^* = 1 + 2x + e$, where x and e are drawn from $N(0, 1)$ and $N(0, .3)$ respectively, and the observed y is obtained as $y = y^*$ for $y^* > 1$ and $y = 0$ for $y^* \leq 1$, i.e., y is censored at 1. The OLS estimate of the coefficient of x (the true value of which is 2) in this case is biased downward, at 1.44. This bias can be seen in Fig. 1 by comparing the slope of the regression line (the longer one) with that of the true line (the shorter one). In Fig. 2,

Table 4 OLS estimation of data censored at various points (True value of coefficient = 1)

Mean estimate	*SD*	*Data censored at*	*% of data uncensored*
.761	.026	.0	.76
.771	.027	.5	.64
.782	.028	1.0	.50
.759	.029	1.5	.36
.679	.033	2.0	.24

[10]The percentages of data (un)censored are shown in Table 4, where the figures are of more direct interest.

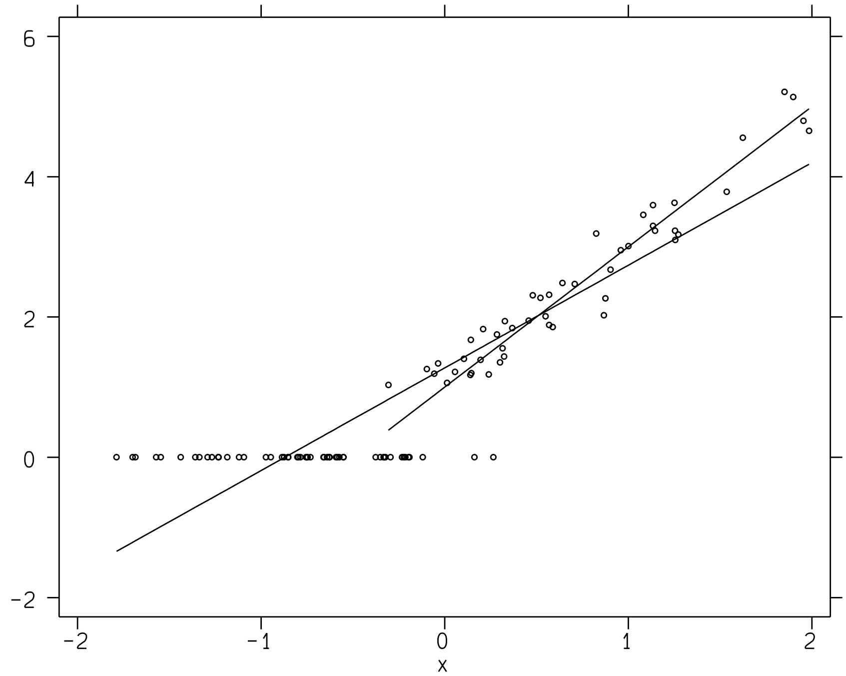

Fig. 1 Downward bias of OLS estimates on censored data.

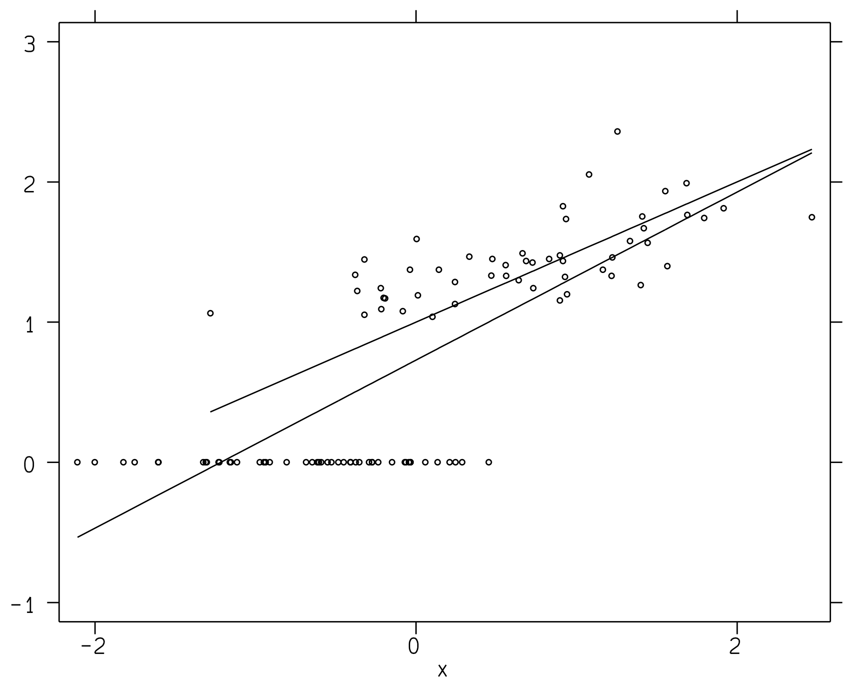

Fig. 2 Upward bias of OLS estimates on censored data.

the latent function is $y^* = 1 + .5x + e$ and the observed y is again derived from y^* with censoring at 1. The OLS estimate of the x coefficient (the true value of which is .5) in this case is biased *upward*, at .53, as the comparison of the two lines reveals.

3 The Heckit Model

Whereas the Tobit model was designed to deal with estimation bias associated with censoring, the Heckit model (Heckman 1979) is a response to sample selection bias, which arises when interest centers on the relationship between $\mathbf{x}$ and y but data are available only for cases in which another variable, z^*, exceeds a certain value.[11] In this model,

$$z_i^* = \mathbf{w}_i\gamma + \mu_i \tag{8}$$

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i \qquad \text{observed only if} \qquad z_i^* > 0 \tag{9}$$

where the error terms are assumed to follow a bivariate normal distribution with means 0, variances[12] $\sigma_\mu = 1$ and σ_ϵ, and correlation coefficient ρ. As in the Tobit case, OLS estimation of model (9) using the observed y is biased. Achen (1986, Chap. 4) provides both an intuitive explanation and formal proofs concerning the nature and magnitude of the bias.

Given the joint distribution, the likelihood of the observed y can be derived and maximum-likelihood estimation can be carried out. However, because maximum-likelihood estimation is computationally cumbersome and sometimes fails to converge for this model, Heckman's (1979) two-step estimator is usually employed instead. The estimator is based on the conditional expectation of the observed y:

$$E(y \mid z^* > 0) = \mathbf{x}\boldsymbol{\beta} + \rho\sigma_\epsilon\lambda(-\mathbf{w}\gamma) \tag{10}$$

where the inverse Mills ratio $\lambda(-\mathbf{w}\gamma) = (\phi(-\mathbf{w}\gamma))/(1 - \Phi(-\mathbf{w}\gamma))$. Equation (10) implies that the conditional expectation of y is $\mathbf{x}\boldsymbol{\beta}$ only when the errors of Eq. (8) and Eq. (9) are uncorrelated; otherwise it is affected by variables in the selection equation as well. The equation also suggests that consistent estimates of $\boldsymbol{\beta}$ can be obtained via OLS regression of the observed y on $\mathbf{x}$ *and* $\lambda(.)$; the unknown coefficient in $\lambda(.)$, γ, can be obtained from a probit estimation of z on $\mathbf{w}$, where $z = 1$ if $z^* > 0$ and 0 otherwise.

In the Heckit model attention usually centers on the observed y, and the marginal effect of the kth element of $\mathbf{x}$ on its conditional expectation is

$$\frac{\partial E(y \mid z^* > 0, \mathbf{x})}{\partial \mathbf{x}_k} = \boldsymbol{\beta}_k - \gamma_k\rho\sigma_\epsilon\delta(-\mathbf{w}\gamma) \tag{11}$$

where the function δ is as defined for Eq. (6) (Greene 1997, p. 977). This shows that the impact of $\mathbf{x}$ is a compound of its impact on the selection and the outcome equations. When the errors in the selection and the y regression equations are correlated ($\rho \neq 0$), it is incorrect to interpret $\boldsymbol{\beta}_k$ as the marginal effect of $\mathbf{x}_k$ on y, unless $\mathbf{x}_k$ does not enter the selection equation (in which case $\boldsymbol{\gamma}_k = 0$).

One limitation of the Heckit model that has received minimal attention is that with the usual operationalization of y as a linear combination of the independent variables, which

[11]Tobit is a special case of the Heckit model, when the selection equation is identical to the regression equation.
[12]The variance of μ is unidentified, so set to 1.

can take on any value, the model is not efficient for exclusively nonnegative data. The model assumes that y is observed whenever another variable, z^*, is positive. This does not restrict the range of values that y itself can take on.[13] Indeed, the nonnegativity of the data is itself explicit information that can and should be used by the model. However, in practice Heckit is inappropriately seized upon as an alternative to Tobit for analyzing exclusively nonnegative data. This results in information loss and inefficient estimation.[14] To correct this problem, the specification of the y equation can be modified by making y itself a Tobit or by using nonnegative functions on y.[15]

In applications of the Heckit model to political data (e.g., Grier et al. 1994; McCarty and Rothenberg 1996), one major problem is the proper interpretation of the model results. Recall that according to Eq. (11), the marginal effect of $\mathbf{x}$ on the observed y is *not* the estimated $\boldsymbol{\beta}$ if $\mathbf{x}$ also enters the selection equation. Rather, that effect is given by $\boldsymbol{\beta}_k - \gamma_k \rho \sigma_\epsilon \delta(-\mathbf{w}\gamma)$. Even so, $\boldsymbol{\beta}$ is often mistakenly interpreted as the marginal effect of $\mathbf{x}$.

We illustrate the correct interpretation of Heckit results by reanalyzing the data Grier et al. (1994) used in their study of corporate PAC donations to congressional candidates. Grier et al. tested a model of corporate PAC donations to House candidates, 1978–1986, aggregating contributions according to the industry to which a firm belonged. For the 620 observations in the data set (124 industries in each of five separate campaigns), contributions ranged from \$0 (the value observed for 216, or 35%, of the cases) all the way to \$2,715,400. The predictors in the model were a set of industry characteristics such as average sales to government by firms in an industry, the geographical concentration of firms in the industry, and the variability of profits over time among industry firms. More specifically, the model was

$$\begin{aligned} y = c &+ \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 x_3 + \beta_5 x_3^2 + \beta_6 x_4 + \beta_7 x_5 + \beta_8 x_6 \\ &+ \beta_9 x_6^2 + \beta_{10} x_7 + \beta_{11} x_8 + \beta_{12} x_9 + \beta_{13} IMR(\mathbf{x}) \end{aligned}$$

where y is the amount contributed by industry firms, x_1 is a time counter tapping a trend across the five campaigns, x_2 is private sales in hundreds of millions of dollars, x_3 is government sales in hundreds of millions of dollars, x_4 is a dummy variable for regulated industries, x_5 is a dummy variable for diverse industries, x_6 is the industrial concentration ratio, x_7 is the number of antitrust indictments against industry firms, x_8 is the variability of profit for industry firms from year to year, x_9 is the geographic concentration of industry firms, and *IMR* is the inverse Mills ratio. Note that nonlinearity enters the tested model not only in the x_2^2, x_3^2, and x_6^2 terms, but also in the inverse Mills ratio, which in this model is a function of all the x variables: $IMR = (\phi(\mathbf{x}\gamma))/(1 - \Phi(\mathbf{x}\gamma))$, where γ is the vector of probit coefficients.

The issue is how to interpret the Heckit estimates. Grier et al. proceeded, as in the OLS context, by multiplying the β for a given x by a specified number of x units and interpreting the product as the change in y associated with the specified amount of change in x. These interpretations are problematic due to the presence of nonlinearity in the *IMR* term. As noted earlier, every predictor in the model appears not only as $x_1 \ldots x_9$ but also as a component

[13] This is in contrast to the Tobit model, which assumes that $y = y^*$ only when $y^* > 0$; otherwise $y = 0$. Hence in the Tobit model, observed y values can never be negative.

[14] And if other assumptions of the model are also violated, e.g., if the underlying data-generating process is Poisson rather than anything like a selection mechanism, then the consequences of applying the Heckit model can be even more serious.

[15] The first alternative is implemented in *LIMDEP* (Greene 1995) as Tobit with selection bias.

of the *IMR*. One consequence of this nonlinearity is that the effect of n units of change in x is not simply n times the effect of one unit of change in x. Another consequence is that the effect of a change in x depends not only on the magnitude of the change, but also on the base from which the change takes place. As Eq. (11) makes clear, a change in an x variable has a compound effect, through both the selection and the contribution equations, and may, for example, not only cause those PACs already giving money to give more money, but also draw in some PACs that have not previously contributed. Thus, to estimate the yield in contributions produced by, say, a unit change on x_7, one cannot simply multiply β_{10} by one. Rather, one must also make corresponding adjustments to the $\beta_{13} IMR(\mathbf{x})$ product.

The Heckit marginal effect formula (11) takes into consideration the effect of the *IMR* term, but is inapplicable here because the tested model specifies nonlinear effects for x_2, x_3, and x_6, whereas the formula is derived for a linear $\mathbf{x}\boldsymbol{\beta}$ specification in the model for y. The correct marginal effects for the tested model can be derived for this specific functional form, which would involve taking partial derivatives of $E(y \mid z^* > 0)$ with respect to the various x. For this particular data set, however, marginal effects would not always be meaningful, because the model contains dummy variables and some other variables, e.g., x_1, x_4, x_5, and x_7, for which the meaning of "small change" is ambiguous. We therefore conducted numerical sensitivity analyses to assess the change in $E(y \mid z^* > 0)$ given a specified change in an x variable.

Our re-estimates of the dollar yields of increments in the variables in the Grier et al. model are not constant across observations, due to the nonlinear *IMR* term in the model. Rather, there is a distinct estimate for each observation, so we have taken the mean of all these distinct estimates as our yield measure. In general, our reestimates are consistently lower than those calculated by Grier et al., and for some variables the difference is very substantial.[16] To illustrate, consider two key variables:

- Grier et al. estimated the impact of \$50 million in private sales (x_2) as \$10600 in PAC contributions, but our re-estimate is \$3804, just 36% of the yield they project. Their overestimate of the impact of private sales leads them to understate the differential effects of private and public sales—a key contrast in their model, which revolves around the idea that industries contribute as a function of the benefits they stand to receive from government assistance. According to Grier et al., the public-private sales yield ratio for \$50 million dollars in sales is approximately 6:1 (\$65500 to \$10600), but according to our re-estimate it is closer to 12:1 (\$45583 to \$3804).
- The Grier et al. estimate of the increased yield in PAC contributions associated with a one-standard deviation increase in the proportion of sales accounted for by the four largest firms in an industry (x_6) is seven times as large as our re-estimate (\$61733 vs. \$8501). Thus, industrial concentration appears to have a much more limited impact on PAC contributions when the x_6 coefficient is appropriately interpreted.

Because our sensitivity estimate for a particular x is actually a mean, aggregated from the distinct sensitivity value for each case, it conveys the "average" impact of an independent variable. However, a more disaggregated approach can aid interpretation, providing information that would be lost if attention were confined to central tendencies. To illustrate, we turn to a key variable in the Grier et al. analysis, government sales (x_3). Across the 404 observations, the expected yield in industry PAC contributions associated with a \$50 million increase in government sales ranges all the way from −\$45,895 to +\$65,877. A

[16] Detailed results can be obtained at the Political Analysis website.

P1: FBI/FKP P2: FBW
WV002-05 December 16, 1999 11:36

logical follow-up question is why industries differ so markedly in the impact of government sales on PAC contributions.

To address this question, we can compare the characteristics of cases with very low sensitivity values on x_3 to those of cases with very high sensitivity values on x_3.[17] It turns out that compared to industries in which increased sales to government strongly depress PAC contributions, industries in which increased government sales produce the greatest yield in contributions give much less in the first place (y), average hundreds of millions of dollars more in private sales (x_2) and hundreds of millions of dollars less in government sales (x_3) (indeed, make no sales at all to government), are less diversified in terms of product lines (x_5), have accumulated more antitrust indictments over the years (x_7), and are much more stable from year to year in terms of profitability (x_8). The opportunity to carry the interpretation forward in this manner would not even arise if, inappropriately, the Heckit coefficients were interpreted like OLS coefficients. In that case, the sensitivity measure for x_3 would take on exactly the same value across all observations instead of varying from observation to observation.

4 Conclusion

We have considered the Tobit and Heckit models for the analysis of censored and sample-selected data. The Tobit model was designed to deal with biases introduced by censoring. When applied under the specific conditions for which it is appropriate, the Tobit model provides a useful supplement to the standard repertoire of models. Problems arise, however, when the standard Tobit model is applied to data for which it is inappropriate, as often occurs. We have pointed to some obvious problems in Tobit-based analyses, and through reanalysis of the Romer–Snyder data as well as numerical experiments, we have demonstrated that when the data come from a generating process other than censoring or when they have been censored but not at zero, the standard Tobit model can produce a poor fit to the data and can seriously bias parameter estimates. As a side note, we have also pointed out that when censoring is not at zero, the well-known proof concerning the direction and degree of bias of OLS estimates on standard Tobit data no longer holds.

The Heckit model, which contains the Tobit model as a special case, was designed as a corrective for sample selection bias. We have pointed out that applying the Heckit model to exclusively nonnegative data, as is often done in analyses of political data, results in information loss and inefficient estimation. We have dealt at greater length with the proper interpretation of Tobit and Heckit estimation results, cataloging problems that ensue when the estimates are interpreted in the same way as OLS estimates. Using Grier et al.'s Heckit model of corporate PAC donations, we have reported sensitivity analyses of the Heckit estimation results. These results suggest some rather different conclusions than Grier et al. reached about the impact of key variables in their model.

As we have moved from the Tobit model to the Heckit model, we have moved from a more restricted model to one of more general applicability. Of course, applying a model to data that violate the model's main assumptions results in incorrect inferences. It does not follow, however, that the more general model is invariably preferable to the more specific one. Indeed, if the more specific model is the correct model for the data, then it is preferable to the more general model; the latter will be inefficient and will result in

[17] For purposes of illustration, we compare the cases with the five highest and the five lowest (in fact, negative) sensitivity values. One could just as easily compare cases above the mean or median with cases below the mean or median, cases ± 1 SD from the mean, or cases selected according to any other criterion.

less powerful statistical tests. In practice, analysts of political data rarely have specific information about the data-generating process. An appropriate analytic strategy, then, is to experiment with more general specifications before reaching substantive conclusions based on a relatively specific and restricted model. Inconsistencies between the results produced by the more limited and the more general model may have major implications for the substantive conclusions one draws.

References

Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Amemiya, Takeshi. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 84:3–61.

Beck, Nathaniel, and Richard Tucker. 1998. "Democracy and Peace: General Law or Limited Phenomenon?" Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Beck, Nathaniel, Gary King, and Langche Zeng. 1998. The Problem with Quantitative Studies of International Conflict. Paper presented at the Political Methodology Summer Meeting, UC San Diego.

Box-Steffensmeier, Janet, and Jones, Bradford. 1997. "Time Is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41:1414.

Breen, Richard. 1998. *Regression Models: Censored, Sample Selected, or Truncated Data*. Quantitative Applications, No. 111. Beverly Hills, CA: Sage.

Chappell, Henry W., Jr. 1982. "Campaign Contributions and Congressional Voting: A Simultaneous Probit-Tobit Model." *Review of Economics and Statistics* 62:77–83.

Dubin, Jeffrey A., and Douglas Rivers. 1989–1990. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods and Research* 18:360–390.

Goldberger, Arthur S. 1972. "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations," Institute for Research on Poverty Discussion Paper 123-72. Madison: University of Wisconsin.

Greene, William. 1981. "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model." *Econometrica* 49:505–513.

Greene, William. 1995. *LIMDEP, Version 7.0.* Bellport, NY: Econometric Software.

Greene, William. 1997. *Econometric Analysis*, 3rd ed. New York: Macmillan.

Grier, Kevin B., and Michael C. Munger. 1993. "Comparing Interest Group PAC Contributions to House and Senate Incumbents, 1980–1986." *Journal of Politics* 55:615–643.

Grier, Kevin B., Michael C. Munger, and Brian E. Roberts. 1994. "The Determinants of Industry Political Activity, 1978–1986." *American Political Science Review* 88:911–926.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.

Herron, Michael C. 1998. "The Presidential Election of 1988: Low Voter Turnout and the Defeat of Michael Dukakis," Unpublished paper.

Jacobs, David, and O'Brien, Robert M. 1998. "The Determinants of Deadly Force: A Structural Analysis of Police Violence." *American Journal of Sociology* 103:837–862.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* New York: Cambridge University Press.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 1998. "Listwise Deletion is Evil: What to Do About Missing Data in Political Science." Paper presented at the Political Methodology Summer Meeting, UC San Diego.

King, Gary, Michael Tomz, and Jason Wittenberg. 1998. "How to Intepret and Present Statistical Results *or* Enough with the Logit Coefficients Already!" Paper prepared for the annual meetings of the American Political Science Association.

Laitila, Thomas. 1993. "A Pseudo-R^2 Measure for Limited and Qualitative Dependent Variable Models." *Journal of Econometrics* 56:341–356.

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

Maddala, G. S. 1992. *Introduction to Econometrics*, 2nd ed. New York: Macmillan.

McCarty, Nolan, and Lawrence S. Rothenberg. 1996. "Commitment and the Campaign Contribution Contract." *American Journal of Political Science* 40:872–904.

McDonald, John F., and Robert A. Moffitt. 1980. "The Uses of Tobit Analysis." *Review of Economics and Statistics* 62:318–321.

Merrnik, James. 1994. “Presidential Decision Making and the Political Use of Military Force.” *International Studies Quarterly* 38:121–138.

Morgan, T. Clifton, and Kenneth N. Bickers. 1992. “Domestic Discontent and the External Use of Force.” *Journal of Conflict Resolution* 36:25–52.

Nagler, Jonathan. 1994. “Scobit: An Alternative Estimator to Logit and Probit.” *American Journal of Political Science* 38:230–255.

Romer, Thomas, and James M. Snyder, Jr. 1994. “An Empirical Investigation of the Dynamics of PAC Contributions.” *American Journal of Political Science* 38:745–769.

Roncek, Dennis W. 1992. “Learning More from Tobit Coefficients: Extending a Comparative Analysis of Political Protest.” *American Sociological Review* 57:503–507.

Tobin, James. 1958. “Estimation of Relationships for Limited Dependent Variables.” *Econometrica* 26:24–36.

Veall, Michael R., and Klaus F. Zimmerman. 1994. “Goodness of Fit Measures in the Tobit Model.” *Oxford Bulletin of Economics and Statistics* 56:485–499.

Walton, John, and Charles Ragin. 1990. “Global and National Sources of Political Protest: Third World Responses to the Debt Crisis.” *American Sociological Review* 55:876–890.

Wang, Kevin H. 1996. “Presidential Responses to Foreign Policy Crises.” *Journal of Conflict Resolution* 40:68–97.