Review of Applied Economics, Vol. 2, No. 1, (2006) : 1-12

CENSORED PROBIT ESTIMATION WITH CORRELATION NEAR THE BOUNDARY: A USEFUL REPARAMETERIZATION

Joseph V. Terza [*] ***and Wei-Der Tsai*** [**]

The conventional computational algorithms for full information maximum likelihood (FIML) estimation of the censored probit model (see Farber, 1983), will sometimes fail to converge when the estimated value of the correlation coefficient (ρ) approaches $\pm$ 1; even when the true value of ρ is not at a boundary. We show that a simple reparametrization of the censored probit model may afford straightforward Newton-Raphson convergence to the true FIML estimate for cases in which likelihood maximization under the conventional censored probit parameterization would have failed. Moreover, our method avoids the computational and inferential complications that arise in alternative methods that, based on a specified criterion, suggest fixing the estimated value of ρ at –1 or +1. For the purpose of illustration the method is used to estimate the determinants of elderly parents' receipt of informal care from their children.

JEL Classifications: *I12, C24, C63*

Keywords: *sample selection, endogenous treatment effects, endogenous switching, qualitative dependent variables, informal elderly care.*

INTRODUCTION

In the present paper we focus on the estimation of a binary response model in which the dependent variable (y) is observable if and only if a binary censoring variable (d) is equal to 1. The probit version of this model, commonly referred to as the *censored probit model*, was first considered by Farber (1983) and has since been widely applied in the literature.[1] In some cases, during iterative full information maximum likelihood (FIML) estimation of this model the estimated value of ρ, the correlation coefficient for y and d (conditional on the exogenous variables), will approach one of its boundaries (+1 or –1) and the algorithm will fail to converge – e.g., the gradient of the likelihood function will not approach zero. We henceforth refer to this phenomenon as the *boundary value problem*. In such situations, one is inclined to terminate the iterations when the correlation estimate $(\hat{\rho})$ is within a prespecified neighborhood of the boundary,

* Department of Epidemiology and Health Policy Research, Department of Economics, University of Florida Gainesville, FL 32610.

** The Graduate Institute of Industrial Economics, National Central University, Chung-Li, 32054, Taiwan, R.O.C.

and take the estimates for the remaining parameters to be those in force at the last iteration. In search of a more rigorous alternative, Butler (1996) derives population conditions that are necessary and sufficient for the population value of ρ to be at a boundary. He suggests that upon termination of the iterations as described above, the sample analogue to these conditions can be checked. If these sample analogue conditions hold, the estimate of ρ is set equal to the relevant boundary and the remaining parameters are estimated via a constrained maximum likelihood (CML) algorithm. Butler (1997) gives Monte Carlo evidence that the boundary value problem does not occur randomly, i.e. when the true value of ρ is not at the boundary (between -.9 and +. 9) the conventional censored probit algorithm will not fail.

The present paper contributes to the discussion by showing that it is possible for the conventional censored probit algorithm to fail due to the boundary value problem when the actual FIML estimate of ρ is not at a boundary. Specifically, we show that a simple reparametrization of the censored probit model may afford straightforward Newton-Raphson convergence to the true FIML estimate for cases in which likelihood maximization under the conventional censored probit parametrization would have failed. This is a noteworthy result because, in such cases, Butler's method mistakenly fixes the estimated value of ρ at –1 or +1. Moreover, our method offers a potentially simple estimation alternative in the presence of the boundary value problem, because it avoids the gradient and Hessian discontinuities that result from fixing ρ at a boundary of the parameter space. Such discontinuities complicate the optimization of Butler's constrained likelihood function and preclude the implementation of standard unconstrained optimization methods like the Newton-Raphson algorithm.

Moreover, because Butler's method involves fixing the value of ρ, asymptotic inference for the full vector of parameters does not incorporate randomness due to the estimation of ρ. Our reparameterized maximum likelihood (RML) algorithm optimizes over all parameters freely and therefore yields standard information matrix based asymptotics in which sampling variation due to all parameter estimates is taken into account. For the purpose of illustration we applied both methods to data graciously supplied by Professor Butler.

In the next section, the conventional censored probit estimator of Farber (1983) is reviewed. In Section 3 Butler's CML estimator is detailed, and our RML approach is introduced in Section 4. In Section 5 we apply our estimator to a real dataset that is plagued by the boundary value problem. For this dataset the conventional censored probit algorithm does not converge ($\hat{\rho}$ approaching –1) and Butler concludes that the FIML estimate of ρ is –1. Our RML algorithm, however, converges when applied to this dataset and yields the true interior FIML estimate of ρ ($\hat{\rho} = -.976$). We have thus demonstrated the possibility that conventional censored probit analysis may fail due to the boundary value problem, when the true FIML estimate of ρ is at an interior point. In such cases Butler's CML method is obviated. This is not to say that our approach will solve the boundary value problem in all cases. There may be empirical contexts in which the population value of ρ is at a boundary and our RML method fails. When this happens Butler's CML method can be attempted. Given the relative computational simplicity of our RML technique however, and the conventionality of its asymptotic properties, we recommend that it be used as the first stage of any estimation protocol designed to deal with the boundary value problem in a censored probit model.

THE CONVENTIONAL CENSORED PROBIT ESTIMATOR[2]

In the conventional censored probit model the value of y, the binary outcome variable of interest, is determined by

$$y = I\,(x\beta + \varepsilon > 0) \tag{1}$$

where x denotes a row vector of observable exogenous variables, β is the conformable column vector of unknown parameters, ε is the unobservable random error term, and I(A) denotes the indicator function which is equal to 1 if condition A holds and 0 otherwise. Values of y are observed if and only if d = 1 where

$$d = I(z\alpha + \nu > 0) \tag{2}$$

z denotes a row vector of observable exogenous variables, α is the conformable column vector of unknown parameters, and ν is the unobservable random error term. In the probit version of the model, first considered by Farber (1983), [ε ν | w] is assumed to be standard bivariate normally distributed with mean vector zero and correlation coefficient ρ, where w denotes the vector comprised of the union of the elements of x and z. In this model, conditional on the exogenous variables (w), the probabilities corresponding to the observable (y, d) pairs are:

$$\Pr\{y = 1, d = 1\} = \Phi(x\beta, z\alpha; \rho)$$

$$\Pr\{y = 0, d = 1\} = \Phi(z\alpha) - \Phi(x\beta, z\alpha; \rho)$$

$$\Pr(d = 0\} = \Pr\{y = 1, d = 0\} + \Pr\{y = 0, d = 0\} = 1 - \Phi(z\alpha)$$

where $\Phi(\cdot)$ and $\Phi(\cdot,\cdot\,;\cdot)$ denote the univariate and bivariate standard normal cdfs, respectively. Therefore, in the conventional censored probit model the log-likelihood of a random sample of size n is

$$L(\beta,\alpha,\rho \mid w) = \sum_{i\in C}\left\{(1-y_i)\,\ln[\Phi(z_i\alpha) - \Phi(x_i\beta, z_i\alpha;\rho) + y_i\,\ln[\Phi(x_i\beta, z_i\alpha;\rho)]\right\}$$

$$+\sum_{i=1}^{n}(1-d_i)\,\ln[1-\Phi(z_i\alpha)] \tag{3}$$

where C denotes the subset of the sample for which d = 1. Maximization of (3) yields estimates of β, α, and ρ that are consistent, asymptotically normal and efficient.

BUTLER'S CONSTRAINED MAXIMUM LIKELIHOOD ESTIMATOR

Iterative optimization of (3) will often fail to converge as the estimated value of ρ approaches one of its boundaries. Butler (1996) offers a method for applied researchers who encounter this boundary value problem. He shows that ρ = 1 if and only if the joint probability that $x\beta \geq z\alpha$ and y = 0 is zero. Similarly, ρ = –1 if and only if the joint probability that $x\beta \leq -z\alpha$ and y = 1 is zero. Butler suggests that, when confronted with the boundary value problem, the researcher should terminate the likelihood iterations when $\hat{\rho}$ is "close" to +1 (–1), and then check the sample analogs to the aforementioned conditions using the estimates of α and β that were in force upon termination of the iterations – call these estimates $\hat{\alpha}$ and $\hat{\beta}$. For example, if $\hat{\rho}$ approaches +1 check to see if there are any observations for which $x_i\hat{\beta} \geq z_i\hat{\alpha}$ *and* $y_i = 0$, where i denotes the i^{th} sample member. If there are no such observations then fix the value of ρ at 1 and maximize the constrained (ρ = 1) version of the likelihood function. A similar procedure is followed when $\hat{\rho}$ approaches –1. When ρ is constrained to be equal to 1, the probabilities corresponding to the observable (y, d) pairs are:[3]

$$\Pr\{y = 1, d = 1\} = I(x\beta \geq z\alpha)\,\Phi(z\alpha) + [1 - I(x\beta \geq z\alpha)]\Phi(x\beta)$$

$$\Pr\{y = 0, d = 1\} = [1 - I(x\beta \geq z\alpha)][\Phi(z\alpha) - \Phi(x\beta)]$$

$$\Pr\{d = 0\} = \Pr\{y = 1, d = 0\} + \Pr\{y = 0, d = 0\} = 1 - \Phi(z\alpha).$$

Therefore, when ρ is constrained to be equal to 1, the log-likelihood of a random sample of size n

$$L_1(\beta,\alpha,|x,z) = \sum_{i\in C}\Big\{(1-y_i)[1 - I(x_i\beta \geq z_i\alpha)]\ \ln[\Phi(z_i\alpha) - \Phi(x_i\beta)]$$
$$+ y_i\big[I(x_i\beta \geq z_i\alpha)\ \ln[\Phi(z_i\alpha)] + [1 - I(x_i\beta \geq z_i\alpha)]\ \ln[\Phi(x_i\beta)]\big]\Big\}$$
$$+ \sum_{i=1}^{n}(1-d_i)\ \ln[1 - \Phi(z_i\alpha)]. \tag{4}$$

Similarly, when ρ is constrained to be equal to –1, the probabilities corresponding to the observable (y, d) pairs are:

$$\Pr\{y = 1, d = 1\} = I(x\beta > -z\alpha)[\Phi(x\beta) + \Phi(z\alpha) - 1]$$

$$\Pr\{y = 0, d = 1\} = I(x\beta > -z\alpha)[1 - \Phi(x\beta)] + [1 - I(x\beta > -z\alpha)]\Phi(z\alpha)$$

$$\Pr\{d = 0\} = \Pr\{y = 1, d = 0\} + \Pr\{y = 0, d = 0\} = 1 - \Phi(z\alpha).$$

Therefore, when ρ is constrained to be equal to –1, the log-likelihood of a random sample of size n is

$$L_2(\beta,\alpha,|w) = \sum_{i\in C}\Big\{(1-y_i)\big[I(x_i\beta > -z_i\alpha)\ln[1 - \Phi(x_i\beta)] + [1 - I(x_i\beta > -z_i\alpha)]\ln[\Phi(z_i\alpha)]\big]$$
$$+ y_i\big[I(x_i\beta > -z_i\alpha)\ \ln[\Phi(x_i\beta) + \Phi(z_i\alpha) - 1]\big]\Big\}$$
$$+ \sum_{i=1}^{n}(1-d_i)\ \ln[1 - \Phi(z_i\alpha)]. \tag{5}$$

There are a number of problems with Butler's constrained maximum likelihood (CML) approach. First, the criterion for termination of the conventional censored probit iterations is arbitrary. How the chosen termination criterion affects the sample analogs to Butler's conditions is not clear. Secondly, the gradients and Hessians of the constrained log-likelihood functions (4) and (5) are fraught with discontinuities due to the presence of the indicator functions. This complicates the implementation of the method because it precludes standard optimization techniques like the Newton-Raphson algorithm. Thirdly, CML standard errors are biased because the computation of the estimated asymptotic covariance matrix does not take account of the fact that the fixed value of ρ is an estimate and is therefore subject to sampling variation. Lastly, it is possible that the CML method will produce a false boundary estimate for ρ when the true FIML estimate is in the interior of the (-1, 1) interval.

THE REPARAMETERIZED MAXIMUM LIKELIHOOD ESTIMATOR

For censored probit models plagued by the boundary value problem, we suggest FIML estimation of a reparameterized version of the conventional model given in (3). Note that under the assumptions of the model we can write

$$\Pr\{y = 1, d = 1\} = \int_{-z\alpha}^{\infty}\int_{-x\beta}^{\infty} \varphi(\varepsilon, \nu; \rho)\, d\varepsilon\, d\nu$$

$$= \int_{-z\alpha}^{\infty}\left[\int_{-x\beta}^{\infty} g(\varepsilon \mid \nu, w)\, d\varepsilon\right]\varphi(\nu)\, d\nu$$

where $\varphi(\varepsilon, \nu; \rho \mid w)$ denotes the standard bivariate normal density with correlation coefficient ρ, and $g(\varepsilon \mid \nu, w)$ is the density of ε conditional on ν and w. But under the assumptions of the model, $g(\varepsilon \mid \nu, w)$ is the normal density with mean νρ and variance $(1 - \rho^2)$. Therefore, after standardizing $(\varepsilon \mid \nu, w)$ we have

$$\int_{-x\beta}^{\infty} g(\varepsilon \mid \nu, w)\, d\varepsilon = \Phi\left(\frac{x\beta + \nu\rho}{\sqrt{1-\rho^2}}\right)$$

so[4]

$$\Pr\{y = 1, d = 1\} = \int_{-z\alpha}^{\infty} \Phi(x\gamma + \theta\nu)\,\varphi(\nu)\, d\nu \qquad (6)$$

where $\gamma = \beta / \sqrt{1-\rho^2}$, and $\theta = \rho / \sqrt{1-\rho^2}$. From (6) the remaining components of the likelihood function follow as

$$\Pr\{y = 0, d = 1\} = \Phi(z\alpha) - H(x\gamma, z\alpha; \theta)$$

and

$$\Pr\{d = 0\} = 1 - \Phi(z\alpha)$$

where

$$H(x\gamma, z\alpha; \theta) = \int_{-z\alpha}^{\infty} \Phi(x\gamma + \theta\nu)\,\varphi(\nu)d\nu.$$

Therefore, the reparameterized censored probit log-likelihood of a random sample of size n is

$$L_3(\gamma,\alpha,\theta \mid x,z) = \sum_{i\in C}\left\{(1-y_i)\ \ln[\Phi(z_i\alpha) - H(x_i\gamma, z_i\alpha;\theta)] + y_i\ \ln[H(x_i\gamma, z_i\alpha;\theta)]\right\}$$

$$+\sum_{i=1}^{n}(1-d_i)\ \ln[1-\Phi(z_i\alpha)]. \tag{7}$$

This log-likelihood function can be maximized with respect to γ, α, and θ to obtain estimates that are consistent, asymptotically normal and efficient.[5,6] Note that the parameter θ is unrestricted in range. We refer to our estimator as the reparameterized maximum likelihood (RML) estimator. The likelihood function (7) is twice continuously differentiable so that it can be optimized via conventional methods.[7] Note that if $\theta = 0$ then d is not endogenous and a simple binary probit estimator will yield consistent estimates of the elements of γ. The null hypothesis that censoring is exogenous can therefore be tested as $H_o : \theta = 0$ using the RML results.[8] Such a test is not possible in Butler's CML framework.

Estimates of the parameters of the original censored probit model (β and ρ) can be obtained via the following obvious transformations

$$\hat{\rho} = \text{sgn}(\hat{\theta})\sqrt{\frac{\hat{\theta}^2}{1+\hat{\theta}^2}} \tag{8}$$

and

$$\hat{\beta} = (\sqrt{1-\hat{\rho}^2})\,\hat{\gamma} \tag{9}$$

where sgn(q) = 1 (–1) if q is positive (negative). The asymptotic t-statistics corresponding to (8) and (9) can be computed using the δ-method.[9] In examining various real datasets that are plagued by the boundary value problem we have found cases in which our RML estimator produces the true FIML interior estimate ρ, but the CML method leads to a spurious boundary estimate. We discuss one such example in the following section.

AN EXAMPLE: INFORMAL CARE FOR THE ELDERLY

There are applied contexts in which the conventional censored probit model of Farber (1983) fails due to the boundary value problem, but our RML algorithm successfully converges to a maximum at an interior estimate of ρ. In order to demonstrate this fact, we applied the RML estimator to a dataset supplied to us by Professor Butler for which the conventional censored probit estimator failed. The data come from a study of the willingness of offspring to supply informal care to elderly parents.[10,11] For example, how does travel time to an elderly parent's home affect the decision to supply informal care? If one is interested in analyzing such effects for the entire elderly population (and their offspring) then conventional estimates may be biased because willingness to supply informal care is only observed if the elderly person is not in a nursing home, though such willingness may exist amongst the offspring of some nursing home residents. Therefore, relative to the population of interest (viz. *all elderly persons*) the observed sample is censored. If the unobservables that affect the decision to enter a nursing home are correlated with the unobservables determining children's willingness to supply informal care to their elderly parents, then censoring is endogenous and estimates that treat the decision to enter a nursing home as exogenous will likely be biased. In this example the dependent variable y is equal to 1 if an elderly person receives informal care from any of her children, 0 otherwise. The censoring variable d denotes whether (d = 0) or not (d = 1) the elderly person resides in a nursing home. The sample is comprised of 2,530 elderly individuals, of whom 2,383 are not in a nursing home. Definitions and sample means of the variables are given in Table 1.

Butler applied the conventional censored probit estimator and found these data to be plagued by the boundary value problem. He then applied his CML method which produced an estimate

Table 1
Variable Means and Definitions

Variable	*Definition*	*Mean*
	Endogenous Variables	
y	= 1 if receives informal care from children, 0 otherwise.	.23 (n=2,383)
d	= 1 if not living in a nursing home), 0 otherwise.	.94 (n=2,530)
	Exogenous Variables (n=2,530)	
URBAN	= 1 lives in an urban area, 0 otherwise	.40
ADL	number of personal activities of daily living that the individual requires assistance with.	1.6
IADL	number of less personal activities of daily living that the individual requires assistance with.	2.26
DT1	Number of children living at home or less than 31 minutes away.	1.15
DTH	Number of children living 31 to 60 minutes away.	.29
OTHK	Number of children living more than 60 minutes away.	1.06
PFC	Price of formal care per hr.	5.00
BEDRT	Number of nursing home beds per 1,000 persons ≥ age 75.	53.70
HH3	Expected Medicaid subsidy for home health (≥ per week).	.90
SEX	= 1 if male, 0 otherwise.	.31
AGE	Age in years.	77.41
BLACK	= 1 if black, 0 otherwise	.31
SPOUSE	= 1 if living with spouse, 0 otherwise	.39
HOUSE	Housing assets ($).	37967.25
WEALNEW	Nonhousing and social security wealth [$x(.001)].	7.59
DIF	Wealth after nursing home price [$x(.001).	1.18
DISC	Medicaid discount [$x(.0001)].	1.77
WEALBET	Financial income + social security income [$x(.001)]	11.39
NBINCS	Social Security income of spouse ($ per month).	212.48

*WEALNEW, DIF, DISC, WEALBET are all dollar amounts, e.g. the mean of WEALNEW is $7,590 x (.001) = 7.59

of ρ equal to –1. Using our RML estimator we obtained the results for the informal care (y) equation which are given in Table 2. For the purpose of comparison, Butler's CML estimates are also displayed therein. The results for the censoring equation are shown in Table 3. The key observation is that Butler's method yielded a boundary estimate of ρ (-1); whereas our RML approach, which uses a simple Newton-Raphson algorithm, produced an interior estimate (-.976). We have thus demonstrated that it is possible for the conventional censored probit model to fail when the true FIML estimate of ρ is interior to the (-1, 1) interval. Moreover, it is implicit that our RML method avoids the complications (discontinuities) of the CML algorithm which follow from fixing the value of ρ at a boundary value.

Mistakenly fixing the estimate of ρ at a boundary will also cause bias in the estimates of the asymptotic standard errors of the remaining parameters for two reasons. First, if the estimate of ρ is fixed at the wrong value, the rest of the parameter estimates will also be biased. Secondly,

Table 2
Informal Care - Dependent Variable = y (t-statistics in parentheses)

Variable	*CML (Butler)*	*RML*
INTERCEPT	-3.00	-3.00
	(-7.17)	(-7.23)
PFC	-.005	-0.015
	(-0.17)	(-0.58)
HH3	-.002	-0.001
	(-0.26)	(-0.15)
DT1	0.31	0.30
	(13.11)	(15.34)
DTH	0.07	0.08
	(1.57)	(1.67)
OTHK	0.02	0.02
	(1.10)	(0.87)
SPOUSE	-0.67	-0.68
	(-6.60)	(-7.03)
WEALBET	-0.02	-0.02
	(-4.58)	(-5.59)
NBINCS	-.008	-.005
	(-0.59)	(-0.44)
ADL	0.04	0.04
	(1.91)	(1.67)
IADL	0.22	0.22
	(11.51)	(12.17)
SEX	-0.14	-0.10
	(-1.80)	(-1.20)
BLACK	-0.02	-.005
	(-0.29)	(-0.07)
URBAN	-0.04	-0.03
	(-0.57)	(-0.49)
AGE	0.22	0.23
	(4.60)	(4.75)
ρ	-1	-.976
	(—)	(-3.12)

Table 3
Not in a Nursing Home - Dependent Variable = d
(t-statistics in parentheses)

Variable	*CML (Butler)*	*RML*
INTERCEPT	2.56	2.72
	(2.93)	(3.26)
ADL	-0.27	-0.27
	(-9.08)	(-9.96)
SEX	-0.17	-0.21
	(-1.28)	(-1.62)
SPOUSE	0.49	0.52
	(2.50)	(2.79)
AGE	-0.14	-0.15
	(-1.61)	(-1.90)
PFC	-.003	-0.01
	(-0.08)	(-0.28)
DIF	-0.06	-0.06
	(-2.36)	(-2.43)
WEALNEW	0.03	0.03
	(1.38)	(1.44)
DTH	0.03	0.04
	(0.39)	(0.48)
OTHK	0.07	0.08
	(1.80)	(2.15)
HH3	-0.02	-0.02
	(-2.10)	(-2.17)
HOUSE	0.16	0.17
	(5.45)	(5.72)
NBINCS	0.10	0.11
	(1.99)	(2.21)
BLACK	1.08	1.08
	(5.15)	(5.18)
DISC	-0.03	-0.02
	(-0.55)	(-0.46)
BEDRT	-0.03	-0.04
	(-0.98)	(-1.11)
DT1	0.30	0.31
	(5.53)	(5.81)

the computation of the estimated asymptotic covariance matrix of the other parameters does not take account of the sampling variation due to the estimate of ρ. To investigate the consequences of such bias in the example, we computed the mean absolute nominal difference between the CML and RML p-values, for all parameters except ρ. We found that the average difference exceeded 5% (in nominal terms). This level of inaccuracy in the size of an hypothesis test can easily lead to incorrect inference.

DISCUSSION

We offer a simple reparameterization of the conventional censored probit likelihood function, and demonstrate that it is capable of producing the *true interior FIML estimate of* ρ in cases where other more complicated methods cannot. We want, however, to be careful not to overstate our case. We are not suggesting that our approach is a panacea for all applied contexts in which the censored probit model fails due to the boundary value problem. There may be instances in which the *true population value of* ρ is indeed at the boundary, in which case Butler's CML technique would be apropos. The present paper merely suggests that, before resorting to the relatively complicated CML method which, if successful, will produce a boundary estimate of ρ, the researcher should apply our relatively simple RML method which might reveal that the FIML estimate of ρ is actually in the interior of the (-1, 1) interval. In this case, in addition to being less computationally burdensome, the RML approach avoids the inherent biases in the CML standard errors and the potential consequent errors in inference.

Morever, if indeed the population value of ρ is at a boundary, the RML approach will yield evidence of this fact because, as expression (8) implies,

$$\rho \to -1 \Leftrightarrow \theta \to -\infty$$

and

$$\rho \to 1 \Leftrightarrow \theta \to \infty.$$

Therefore, if during the RML iterations the estimated value of $|\theta|$ appears to increase without bound with no evidence that the optimization algorithm is converging, one can terminate the iterations at a prespecified value of θ, check Butler's conditions, and proceed with CML estimation.[12]

NOTES

1. This model is often referred to as the *probit model with sample selection*. For examples of applications of binary response models with sample selection see Abdel-khalik, 1994; Boyes et al., 1989; Butler, 1996; Chaykowski et al., 1992; Dubin and Rivers, 1989 and 1990; Farber, 1983; Farber and White, 1990; Jappelli, 1990; Kenkel and Terza, 1999; Meng and Schmidt, 1985; Rivers and Vuong, 1988; Stanley and Coursey, 1990; Van de Ven and Van Praag, 1981; and Van de Ven and Van Vliet, 1995.
2. Note that the present discussion is placed in the context of the censored (sample selection) model, but it can be easily extended to cover models involving an endogenous treatment effect, and the more general class of endogenous switching models (of which sample selection and treatment effect models are special cases).
3. See Butler (1996) Table 1.
4. A more general from of this result can be found in Zeger et al. (1988). We thank Scott Shonkwiler for leading us to this reference.
5. Although the integral in (6) does not have a closed-from it can be accurately and efficiently evaluated using simple Gaussian quadrature–as is programmed in the INTQUADI procedure in GAUSS® (Aptech Systems Inc., 2005).
6. Alternative estimators such as the conditional maximum likelihood approach of Vuong (1984) are possible, but the FIML approach we suggest: (1) is more efficient; (2) has an objective function that imposes minimal marginal programming burden; and (3) does not require that the standard errors be corrected for the two-stage nature of the estimator.
7. A GAUSS® program implementing the Newton-Raphson optimization of (7) is available from the authors. The optimization algorithm terminates when the sum of the squared gradient elements is

less than 10^{-4}. Starting values are obtained using the consistent but less efficient two-stage method of moments estimator developed by Terza (2006).

8. This test for censoring is the simple Wald test of $\theta = 0$. This of course requires that you first estimate the model under the alternative. A simple Lagrange multiplier (LM) test, requiring estimation only under the null, could be developed be based on the likelihood function in (7).
9. See Greene (2000) p. 118.
10. See Sloan et al. (1992) for details on the policy implications of such analyses.
11. For the purpose of comparison, we applied LIMDEP to these data but the algorithm terminated abnormally. The method used in STATA is the one suggested by Van de Ven and Van Praag (1981), but this method is clearly ad hoc and inconsistent. Therefore, the use of STATA in this context will not provide a fair comparison.
12. The prespecified value of θ would of course be chosen to correspond to a value of ρ that is a small distance away from the relevant boundary.

REFERENCES

Abdel-khalik, A. R. (1994), "Performance Evaluation, Tenure And Compensation of CEOs in Different Labor Markets," Working Paper, University of Florida.

Aptech Systems, Inc. (2005), *GAUSS© Programming Language*, Maple Valley, WA: Aptech Systems, Inc.

Boyes, W. J., Hoffman, D. L., and Low, S. A. (1989), "An Econometric Analysis of the Bank Credit Scoring Problem," *Journal of Econometrics*, 40: 3-14.

Butler, J. S. (1996), "Estimating the Correlation in Censored Probit Models," *Review of Economics and Statistics*, 78: 356-358.

——(1997), "Censored Probit Models Do Not Fail Randomly: A Monte Carlo Study," *Economics Letters*, 57: 33-37.

Chaykowski, R.P., Slotsve, G.A., and Butler, J.S. (1992), "A simultaneous analysis of grievance activity and outcome decisions," *Industrial and Labor Relations Review*, 45: 724-737.

Dubin, J. A. and Rivers, D. (1989), "Selection Bias in Linear Regression, Logit and Probit Models," *Sociological Methods and Research*, 18: 360-390.

——(1990), *Selection Bias in Linear Regression, Logit and Probit Models*, in Fox and Long, *Modern Methods of Data Analysis*, Newbury Park: Sage.

Farber, H. S. (1983), "Worker Preferences for Union Representation," *Research in Labor Economics, Supplement 2*: 171-205.

——and White, M. J. (1990), "Medical Malpractice—An Empirical Examination of the Litigation Process," Working Paper, Center for Advanced Study in the Behavioral Sciences, University of Michigan.

Greene, W.H. (2000), *Econometric Analysis 3 Edition*, New York: Prentice-Hall.

Jappelli, T. (1990), "Who is Credit Constrained in the U.S. Economy?" *Quarterly Journal of Economics*, 420: 219-234.

Kenkel, D. S. and Terza, J. V. (1999), "The Demand for Preventive Medical Care," Working Paper, Department of Economics, Penn State University.

Meng, C. L. and Schmidt, P. (1985), "On the Cost of Partial Observability in the Bivariate Probit Model," *International Economic Review*, 26: 71-83.

Rivers, D. and Vuong, Q. H. (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39: 347-336.

Sloan, F. A., Hoerger, T. J., and Picone, G. (1992), "Effects of Strategic Behavior and Public Subsidies on Family's Savings and Long-Term Decisions," Working Paper, Vanderbilt University.

Stanley, L. R. and Coursey, D. L. (1990), "Empirical Evidence on the Selection Hypothesis and the Decision to Litigate or Settle," *Journal of Legal Studies*, 19: 145-172.

Terza, J.V (2006), "Parametric Nonlinear Regression with Endogenous Switching," Working Paper, Department of Epidemiology and Health Policy Research, University of Florida.

Van de Ven, W. P. M. M. and Van Vliet, R. C. J. A. (1995), "Consumer Information Surplus and Adverse Selection in Competitive Health Insurance Markets: An Empirical Study," *Journal of Health Economics*, 14: 149-169.

Van de Ven, W. P. M. M. and Van Praag, B. M. S. (1981), "The Demand for Deductibles in Private Health Insurance - A Probit Model with Sample Selection," *Journal of Econometrics*, 17: 229-252.

Vuong, Q.H. (1984), "Two-Stage Conditional Maximum Likelihood Estimation of Econometric Models," Working Paper #538, California Institute of Technology.

Zeger, S. L., Liang, K., and Albert, P. S. (1988). "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, 44: 1049-1060.