Fitting and interpreting Cragg’s tobit alternative using Stata

William J. Burke
Department of Agricultural, Food, and Resource Economics
Michigan State University
East Lansing, MI
burkewi2@msu.edu

Abstract. In this article, I introduce the user-written command craggit, which simultaneously fits both tiers of Cragg’s (1971, *Econometrica* 39: 829–844) “two-tier” (sometimes called “two-stage” or “double-hurdle”) alternative to tobit for corner-solution models. A key limitation to the tobit model is that the probability of a positive value and the actual value, given that it is positive, are determined by the same underlying process (i.e., the same parameters). Cragg proposed a more flexible alternative that allows these outcomes to be determined by separate processes through the incorporation of a probit model in the first tier and a truncated normal model in the second. Also, tobit is nested in craggit, making the latter a popular choice among “two-tier” models. In the article, I also present postestimation syntax to facilitate the understanding and interpretation of results.

Keywords: st0179, craggit, two tier, two stage, double hurdle, corner-solution models, Craggit’s tobit

1. Introduction

Introductory econometric texts commonly reference tobit for fitting models with limited dependent variables that are sometimes called “corner-solution” models.\(^1\) Fitting and interpreting the tobit model is fairly straightforward through the use of tobit and associated postestimation commands in Stata.\(^2\) A key limitation to the tobit model is that the probability of a positive value and the actual value, given that it is positive, are determined by the same underlying process (i.e., the same parameters). Cragg (1971) proposed a more flexible alternative, sometimes called a “two-tier” or “double-hurdle” model, which allows these outcomes to be determined by separate processes. Thus far, fitting Cragg’s tobit alternative using Stata has required a disjointed process, and interpreting results has been complicated and tedious.

In this article, I present a command, craggit, that enables a more coherent fitting of Cragg’s model, as well as presenting postestimation syntax to facilitate the interpretation of results. The model in this article is applied to cross-sectional data. The craggit command can be used with either balanced or unbalanced panel data as well.

\(^1\) For examples, see Pindyck and Rubinfeld (1998), Kennedy (2003), Wooldridge (2009), or Baum (2006).

\(^2\) For example, see Baum (2006, 262–266).
but only as a pooled estimator (e.g., the `vce(cluster)` option can be used to compute standard errors robust to autocorrelation, but the command is not designed to control for unobserved heterogeneity).

Conceptually, a corner-solution model is where

\[ y_i = y_i^* \quad \text{if} \quad y_i^* > 0 \]
\[ y_i = 0 \quad \text{if} \quad y_i^* \leq 0 \]

and

\[ y_i^* = \alpha + X_i \beta + \varepsilon_i \]

In practice, as the name suggests, a corner-solution model applies to dependent variables where data are truncated and “piles up” at some given value, but is continuous otherwise. Some examples would be agricultural production or input demands where some observations’ optimizing behavior results in no production or demand, while the outcome is continuous for others.

2 The tobit model

Introduced by Tobin (1958), the tobit model proposes the likelihood function,

\[ f(y \mid x_1) = \{1 - \Phi(x_1 \beta / \sigma)\}^{1(y=0)} \left[ (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp \left\{ -(y - x_1 \beta)^2 / 2\sigma^2 \right\} \right]^{1(y>0)} \]

where \( \Phi \) is the standard normal cumulative distribution function and exponential indicator functions are \( 1(y=0) \) and \( 1(y>0) \). There are four values of interest after fitting a tobit model: 1) the probability that \( y \) is zero, \( P(y_i = 0 \mid x_i) \); 2) the probability that \( y \) is positive, \( P(y_i > 0 \mid x_i) \); 3) the expected value of \( y_i \), conditional on \( y_i \) being positive, \( E(y_i \mid y_i > 0, x_i) \); and 4) the so-called “unconditional” expected value of \( y_i \), \( E(y_i \mid x_i) \).

Then, of course, one could calculate the marginal effects of an explanatory variable on each probability and expectation. Fitting this model is fairly simple using the `tobit` command in Stata, and calculation of these effects around data mean values can be obtained using various `margins` postestimation commands. For a thorough discussion on the tobit model and its interpretation, refer to Wooldridge (2009, 587–595).

3 Cragg’s alternative to the tobit model

Again, while useful, the major drawback of the tobit model is that the choice of \( y > 0 \) and the value of \( y \), given that \( y > 0 \), is determined by the same vector of parameters (\( \beta \) from above). For example, this imposes that the direction (sign) of a given determinant’s marginal effect will be the same on both the probability that \( y > 0 \) and the expectation of \( y_i \), conditional or otherwise. As an alternative, Cragg proposed the following, which

3. The commonly used phrase “unconditional expectation” is a bit of a misnomer because all expectations are conditional on the explanatory variables (\( x \)).
Cragg's tobit alternative using Stata

integrates the probit model to determine the probability of \( y > 0 \) and the truncated normal model for given positive values of \( y \),

\[
f(w, y | x_1, x_2) = \{1 - \Phi(x_1 \gamma)\}^{1(w=0)} \left[ \Phi(x_1 \gamma)(2\pi)^{-\frac{1}{2}} \sigma^{-1} \right] \exp \left\{ -\frac{(y - x_2 \beta)^2}{2\sigma^2} / \Phi(x_2 \beta / \sigma) \right\}^{1(w=1)}
\]

where \( w \) is a binary indicator equal to 1 if \( y \) is positive and 0 otherwise. Notice in Cragg’s model the probability of \( y > 0 \) and the value of \( y \), given \( y > 0 \), are now determined by different mechanisms (the vectors \( \gamma \) and \( \beta \), respectively). Furthermore, there are no restrictions on the elements of \( x_1 \) and \( x_2 \), implying that each decision may even be determined by a different vector of explanatory variables altogether. Also notice that the tobit model is nested within Cragg’s alternative because if \( x_1 = x_2 \) and \( \gamma = \beta / \sigma \), the models become identical. For a more thorough discussion of this and other double-hurdle alternatives to tobit, refer to Wooldridge (2002, 536–538).

Fitting Cragg’s alternative requires the additional assumption of conditional independence for the latent variable’s distribution, or \( D(y^* | w, x) = D(y^* | x) \).\(^4\) Please refer to Wooldridge (forthcoming) for a thorough discussion. In short, this implies that the unbiasedness of results is sensitive to model misspecification.

From Cragg’s model, we can obtain the same probabilities and expected values as with tobit by using an updated functional form. The probabilities regarding whether \( y \) is positive are

\[
P(y_i = 0 | x_{1i}) = 1 - \Phi(x_{1i} \gamma)
\]

\[
P(y_i > 0 | x_{1i}) = \Phi(x_{1i} \gamma)
\]

The expected value of \( y \), conditional on \( y > 0 \) is

\[
E(y_i | y_i > 0, x_{2i}) = x_{2i} \beta + \sigma \times \lambda(x_{2i} \beta / \sigma)
\]

where \( \lambda(c) \) is the inverse Mills ratio (IMR)

\[
\lambda(c) = \frac{\phi(c)}{\Phi(c)}
\]

where \( \phi \) is the standard normal probability distribution function. Finally, the “unconditional” expected value of \( y \) is

\[
E(y_i | x_{1i}, x_{2i}) = \Phi(x_{1i} \gamma) \{x_{2i} \beta + \sigma \times \lambda(x_{2i} \beta / \sigma)\}
\]

For a given observation, the partial effect of an independent variable, \( x_j \), around the probability that \( y > 0 \) is\(^5\)

\[
\frac{\partial P(y > 0 | x_1)}{\partial x_j} = \gamma_j \phi(x_1 \gamma)
\]

\(^4\) As described in Wooldridge (forthcoming), the exponential type-two tobit relaxes this assumption; however, convergence can be difficult, especially when \( x_1 = x_2 \).

\(^5\) The marginal effect around the probability \( y = 0 \) is the negative of the value in \(^5\).
where $\gamma_j$ is the element of $\gamma$ representing the coefficient on $x_j$. Equations (1), (2), and (5) are the same as the probabilities and partial effect from a probit regression of $w$ on $x_1$. The partial effect of an independent $x_j$ on the expected value of $y$, given $y > 0$, is

$$\frac{\partial E(y_i \mid y_i > 0, x_{2i})}{\partial x_j} = \beta_j \{1 - \lambda(x_2\beta/\sigma + \lambda(x_2\beta/\sigma))\}$$

(6)

where $\beta_j$ is the element of $\beta$ representing the coefficient on $x_j$. Equations (3) and (6) are the same as the expected values and partial effect from a truncated normal regression of $y$ on $x_2$, with emphasis that the effect is conditional on $y$ being positive.

The partial effect of an independent $x_j$ on the “unconditional” expected value of $y$ is somewhat trickier, because it depends on whether $x_j$ is an element of $x_1$, $x_2$, or both. First, if $x_j$ is an element of both vectors, the partial effect is

$$\frac{\partial E(y \mid x_1, x_2)}{\partial x_j} = \gamma_j \phi(x_1, \gamma) \times \{x_2\beta + \sigma \times \lambda(x_2\beta/\sigma)\}$$

$$+ \Phi(x_1, \gamma) \times \beta_j \{1 - \lambda(x_2\beta/\sigma + \lambda(x_2\beta/\sigma))\} \text{ if } x_j \in x_1, x_2$$

(7)

Now, if $x_j$ is only determining the probability of $y > 0$, then $\beta_j = 0$, and the second term on the right-hand side of (7) is canceled. On the other hand, if $x_j$ is only determining the value of $y$, given that $y > 0$, then $\gamma_j = 0$, and the first right-hand side term in (7) is canceled. In either of the latter cases, the marginal effect will still be a function of parameters and explanatory variables in both tiers of the regression.

4 Fitting the Cragg model using Stata

The maximum likelihood estimation (MLE) of $\gamma$ can be obtained by regressing $w$ on $x_1$ by using probit in Stata. Similarly, the MLE of $\beta$ and $\sigma$ can be obtained by regressing $y$ on $x_2$ by using truncreg. Or, all parameters can be fit simultaneously using craggit.

The syntax is

```
craggit depvar1 [indepvars1] [if] [in] [weight], second(depvar2 [indepvars2]) noconstant vce(vcetype) hetero(varlist) level(#) maximize_options
```

`depvar1` is the indicator variable for whether $y > 0$ ($w$ in the above notation), and `indepvars1` are the independent variables explaining that decision ($x_1$ in the above notation). `depvar2` is the continuous response for positive values ($y$ from above), and `indepvars2` are the independent variables explaining it ($x_2$ from above). `second()` is so named because the determinants of $y$, given $y > 0$, are traditionally thought of as the second “tier” or “hurdle” of the model.
Cragg’s tobit alternative using Stata

Using an example of household-level demand for subsidized fertilizer in Zambia (these data are described in the appendix available online⁶), output from craggit estimation is presented as follows:

```
. craggit basal_g disttown cland educ age,
    > second(qbasal_g disttown cland educ age)
Estimating Cragg’s tobit alternative
Assumes conditional independence
initial: log likelihood = -\infty \quad \text{(could not be evaluated)}
feasible: log likelihood = -1.316e+08
rescale: log likelihood = -692711.38
rescale eq: log likelihood = -7868.8964
Iteration 0: log likelihood = -7868.8964
(output omitted)
Iteration 12: log likelihood = -7498.6146

Number of obs = 6378
Wald chi2(4) = 310.27
Log likelihood = -7498.6146 Prob > chi2 = 0.0000

| Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-------|-----------|-----|------|----------------------|
| Tier1 |           |     |      |                      |
| disttown | -.0069824 | .0010228 | -6.83 | 0.000 | -.008987 \quad -.0049779 |
| cland  | .1013351  | .008374  | 12.10 | 0.000 | .0849223 \quad .1177478 |
| educ   | .0510291  | .008374  | 12.10 | 0.000 | .0349223 \quad .0671478 |
| age    | .066328    | .0041688 | 16.14 | 0.000 | .0581641 \quad .0744916 |
| _cons  | -1.717704  | .0929843 | -18.47| 0.000 | -1.89995 \quad -1.535458 |
| Tier2  |           |     |      |                      |
| disttown | -4.082447 | 12.82117 | -0.32 | 0.750 | -29.21148 \quad 21.04659 |
| cland  | 346.8557   | 195.8594 | 1.77  | 0.077 | -37.02145 \quad 730.733 |
| educ   | 248.2076   | 178.2567 | 1.39  | 0.164 | -101.1691 \quad 597.5843 |
| age    | 12.1808    | 20.47023 | 0.60  | 0.549 | -27.94012 \quad 52.30172 |
| _cons  | -10131.47  | 7309.131 | -1.39 | 0.166 | -24457.1 \quad 4194.162 |
| sigma  |           |     |      |                      |
| _cons  | 1110.764   | 391.2037 | 2.84  | 0.005 | 344.0187 \quad 1877.509 |
```

Results from the section labeled Tier1 are the MLE of \( \gamma \), and those labeled Tier2 are the MLE of \( \beta \) from Cragg’s likelihood function. The constant term in the section labeled sigma is the MLE of \( \sigma \).

Whether estimations are obtained simultaneously or one regression at a time, the results will be identical because of the separability of Cragg’s likelihood function. That is, while using craggit makes estimation more coherent, it will not change results. The primary benefit of using craggit is its ability to facilitate postestimation analysis and interpretation.

I emphasize that, as with all double-hurdle models, separability in estimation does not imply separability in interpretation. To illustrate this point, consider a variable that explains the continuous value of \( y \) but not the probability that \( y > 0 \). If fit separately, the

---

the partial effect of such a variable on the unconditional expected value of \( y \) [the second right-hand-side term in (7)] would be a function of probit results, even though it would not be included in the probit regression. For another example, \( x_j \) may be in both explanatory vectors and have countervailing impacts in each tier (i.e., \( \beta_j \) and \( \gamma_j \) may have different signs). Here the direction of the overall effect cannot be known unless all results are considered together, as in (7).

5 Interpreting results after fitting with \texttt{craggit}

Postestimation analysis commands will be presented for a general case, but it may be useful to follow along with the example by using Zambian fertilizer data presented in the online appendix. After estimation using \texttt{craggit}, new variables representing the three scalar values (\( x_1 \hat{\gamma}, x_2 \hat{\beta}, \hat{\sigma} \)) can be generated for each observation using \texttt{predict}:

\begin{verbatim}
predict xig, eq(Tier1)
predict x2b, eq(Tier2)
predict sigma, eq(sigma)
\end{verbatim}

The bold terms in the above syntax are the names of the new variables generated by the \texttt{predict} commands. The choice of variable name is not important to the final results, except that the names should be used consistently throughout. For this article, generated variable names (and later, program names) will continue to be in bold and will endeavor to be consistent with the notation in (1)–(7).

\texttt{craggit} is equipped to handle a model for heteroskedastic standard errors that is a function of observables with the \texttt{hetero(varlist)} option. If that is used, the above command will generate unique values of \( \sigma \) for each observation (\( \hat{\sigma}_i \)).

Now all the information we need to calculate the predicted values in (1)–(4) and partial effects in (5)–(7) is either predicted as a new variable or stored in Stata’s active memory. The following commands will generate new variables with unique values for each observation that can then be aggregated to any desired level.

To calculate the probability that \( y = 0 \) from (1):

\begin{verbatim}
generate Pw0 = 1 - normal(xig)
\end{verbatim}

To calculate the probability that \( y > 0 \) from (2):

\begin{verbatim}
generate Pw1 = normal(xig)
\end{verbatim}

To calculate expected values, it is useful to generate a new variable for the IMR:

\begin{verbatim}
generate IMR = normden(x2b/sigma)/normal(x2b/sigma)
\end{verbatim}

To calculate the expected value of \( y \), given that \( y > 0 \), from (3):

\begin{verbatim}
generate Eyx2 = x2b + sigma*IMR
\end{verbatim}
Cragg's tobit alternative using Stata

To calculate the unconditional expected value of \( y \) from (4):

\[
\text{generate } \text{Ey}x12 = \text{normal}(x1g) \ast (x2b + \text{sigma}\ast\text{IMR})
\]

If \( \text{indepj} \) is an independent variable of interest (\( x_j \) from the above notation), then to calculate the partial effect on the probability that \( y > 0 \) from (5):

\[
\text{generate } \text{dPixi} = \text{[Tier1]}_{b}\text{[indepj]}\ast\text{normalden}(x1g)
\]

To calculate the partial effect on the expected value of \( y \), given \( y > 0 \) from (6):

\[
\text{generate } \text{dEy}x2 = \text{[Tier1]}_{b}\text{[indepj]}\ast(1-\text{IMR}\ast(x2b/\text{sigma}\ast\text{IMR}))
\]

Finally, to calculate the partial effect on the unconditional expected value of \( y \) from (7):

\[
\text{generate } \text{dEy} = \text{[Tier1]}_{b}\text{[indepj]}\ast\text{normalden}(x1g)\ast(x2b+\text{sigma}\ast\text{IMR})
\]

Once these predicted values have been generated for each observation, the average partial effect (APE) of the independent variable can be found using the \text{summarize} command. For example, the APE on the unconditional expected value of \( y \) is the mean of \( \text{dEy} \):

\[
\text{summarize } \text{dEy}
\]

We can also see how this APE varies across groups by using the \text{tabulate} command. Suppose we have a categorical variable, \text{catvar}:

\[
\text{tabulate catvar, summarize(dEy)}
\]

The standard deviations reported by these summaries describe only the data and should not be considered a parameter estimate. That is, the standard deviation of the predicted partial effects should not be used as a standard error (SE) for inference on the APE. One viable option for inference on an APE is bootstrapping. Bootstrapping starts by reestimating the model and generating a new APE on a random subsample within the data. This process is repeated many times until many APEs have been computed from numerous random subsamples. The standard deviation from those APEs can be used as an SE for the full sample APE.

Fortunately, Stata can be programmed to run this process for you, albeit after several lines of programming:\footnote{Thank you to David Tschirley and Ana Fernandez for providing syntax to bootstrap APE standard errors.}
Here, again, the variable and program names in bold type can be chosen subjectively, so long as they are used consistently and none of the variables generated by the program has the same name as the existing variables. Again the explanatory variable whose APE is being computed is \textit{indepj}. The \# sign must be replaced by the number of times Stata should iterate the bootstrap process (the more iterations there are, the longer the process will take, at about 5–10 seconds per iteration; 100 is a reasonable starting point). While the bootstrap is running, after the last command from above is entered, Stata provides progress updates:

\textbf{Bootstrap replications (100)}

\begin{verbatim}
.x........x..x.x..............................x... 100
.x........x.....................................x... 50
.x........x.................................2
.x........x..................................3
.x........x..................................4
.x........x..................................5
\end{verbatim}

where each “.” represents a completed iteration and the occasional “x” represents a “failed” iteration. Here the failed iterations indicate that Stata was not able to fit coefficients for that subsample. This is likely because the random draw did not provide enough variation in the binary indicator variable (\textit{w}) for \textit{craggit} to converge. Stata ignores these failures, and they will not have much impact on the final results, except to reduce the number of iterations used to compute the bootstrap SE.

To calculate the APE for a different explanatory variable, simply rewrite all the commands from \textit{program} to \textit{bootstrap}, changing the name of \textit{indepj}. First, however, the original program must be cleared from Stata’s memory:

\begin{verbatim}
capture program drop /BT/C8/BX/CQ/D3/D3/D8
cap
\end{verbatim}

Unfortunately, bootstrapping can be a time-consuming process, especially when conducting inference on the APE of multiple variables. An alternative would be to approximate a standard error using the delta method. This is when an SE is approximated using a Taylor expansion around the data mean, which can be accomplished after estimation with \textit{craggit} by using the \textit{nlcom} postestimation command and a few local macros.\footnote{Thank you to Joleen Hadrich and Joshua Ariga for assistance on the syntax for the delta method inference.} After fitting and predicting \textit{xig} and \textit{x2b}:
Summarize X1g
local Xig = r(mean)
Summarize X2b
local X2bb = r(mean)
Nlcom [Tier1]_b[indepj]*normalden(Xig)*X2bb*([sigma]_b[cons]* ///
(normalden(X2bb/[sigma]_b[cons])/normal(X2bb) ///
([sigma]_b[cons]))*[Tier2]_b[indepj]*normal(Xig) ///
*(1-normalden(X2bb/[sigma]_b[cons])/normal(X2bb) ///
([sigma]_b[cons]))*([sigma]_b[cons]+normalden(X2bb/[sigma]_b[cons])
(normal(X2bb/[sigma]_b[cons])))

The SE provided after nlcom can be used with the APE from the summarize command
to compute the p-value manually\(^9\) (the p-value provided after nlcom is only valid for
the partial effect at the mean of \(x\), not the APE).

6 References
Baum, C. F. 2006. An Introduction to Modern Econometrics Using Stata. College
Station, TX: Stata Press.

Cragg, J. G. 1971. Some statistical models for limited dependent variables with application


Wooldridge, J. M. 2002. Econometric Analysis of Cross Section and Panel Data. Cam-
bridge, MA: MIT Press.


———. Forthcoming. Econometric Analysis of Cross Section and Panel Data. 2nd ed.
Cambridge, MA: MIT Press.

About the author
William J. Burke is a PhD student and agricultural economics research specialist at Michigan
State University. He has coauthored several policy analysis studies focused on Africa and has
a strong interest in facilitating the use of nonstandard model structures.

---

\(^9\) See page 5 of the online appendix for an example of computing the p-value manually.