HURDLE AND “SELECTION” MODELS

Jeff Wooldridge
Michigan State University
BGSE/IZA Course in Microeconometrics
July 2009

1. Introduction
2. A General Formulation
3. Truncated Normal Hurdle Model
4. Lognormal Hurdle Model
5. Exponential Type II Tobit Model

1. Introduction

- Consider the case with a corner at zero and a continuous distribution for strictly positive values.
- Note that there is only one response here, y, and it is always observed. The value zero is not arbitrary; it is observed data (labor supply, charitable contributions).
- Often see discussions of the "selection" problem with corner solution outcomes, but this is usually not appropriate.

- Example: Family charitable contributions. A zero is a zero, and then we see a range of positive values. We want to choose sensible, flexible models for such a variable.
- Often define w as the binary variable equal to one if $y > 0$, zero if $y = 0$. w is a deterministic function of y. We cannot think of a counterfactual for y in the two different states. ("How much would the family contribute to charity if it contributes nothing to charity?" "How much would a woman work if she is out of the workforce?")

- Contrast previous examples with sound counterfactuals: How much would a worker in if he/she participates in job training versus when he/she does not? What is a student's test score if he/she attends Catholic school compared with if he/she does not?
- The *statistical* structure of a particular two-part model, and the Heckman selection model, are very similar.

- Why should we move beyond Tobit? It can be too restrictive because a single mechanism governs the "participation decision" ($y = 0$ versus $y > 0$) and the "amount decision" (how much y is if it is positive).
- Recall that, in a Tobit model, for a continuous variable x_j, the partial effects on $P(y > 0|\mathbf{x})$ and $E(y|\mathbf{x}, y > 0)$ have the same signs (different multiples of β_j). So, it is impossible for x_j to have a positive effect on $P(y > 0|\mathbf{x})$ and a negative effect on $E(y|\mathbf{x}, y > 0)$. A similar comment holds for discrete covariates.

- Furthermore, for continuous variables x_j and x_h,

$$\frac{\partial P(y > 0|\mathbf{x})/\partial x_j}{\partial P(y > 0|\mathbf{x})/\partial x_h} = \frac{\beta_j}{\beta_h} = \frac{\partial E(y|\mathbf{x}, y > 0)/\partial x_j}{\partial E(y|\mathbf{x}, y > 0)/\partial x_h}$$

- So, if x_j has twice the effect as x_h on the participation decision, x_j must have twice the effect on the amount decision, too.
- Two-part models allow different mechanisms for the participation and amount decisions. Often, the economic argument centers around fixed costs from participating in an activity. (For example, labor supply.)

2. A General Formulation

- Useful to have a general way to think about two-part models without specif distributions. Let w be a binary variable that determines whether y is zero or strictly positive. Let y^* be a nonnegative, continuous random variable. Assume y is generated as

$$y = w \cdot y^*.$$

- Other than w being binary and y^* being continuous, there is another important difference between w and y^*: we effectively observe w because w is observationally equivalent to the indicator $1[y > 0]$ ($P(y^* = 0)$). But y^* is only observed when $w = 1$, in which case $y^* = y$.

- Generally, we might want to allow w and y^* to be dependent, but that is not as easy as it seems. A useful assumption is that w and y^* are independent conditional on explanatory variables $\mathbf{x}$, which we can write as

$$D(y^*|w, \mathbf{x}) = D(y^*|\mathbf{x}).$$

- This assumption typically underlies *two-part* or *hurdle* models.
- One implication is that the expected value of y conditional on $\mathbf{x}$ and w is easy to obtain:

$$E(y|\mathbf{x}, w) = w \cdot E(y^*|\mathbf{x}, w) = w \cdot E(y^*|\mathbf{x}).$$

- Sufficient is conditional mean independence,

$$E(y^*|\mathbf{x}, w) = E(y^*|\mathbf{x}).$$

- When $w = 1$, we can write

$$E(y|\mathbf{x}, y > 0) = E(y^*|\mathbf{x}),$$

so that the so-called "conditional" expectation of y (where we condition on $y > 0$) is just the expected value of y^* (conditional on $\mathbf{x}$).

- The so-called "unconditional" expectation is

$$E(y|\mathbf{x}) = E(w|x)E(y^*|\mathbf{x}) = P(w = 1|\mathbf{x})E(y^*|\mathbf{x}).$$

- A different class of models explicitly allows correlation between the participation and amount decisions Unfortunately, called a *selection model*. Has led to considerable conclusion for corner solution responses.
- Must keep in mind that we only observe one variable, y (along with $\mathbf{x}$). In true sample selection environments, the outcome of the selection variable (w in the current notation) does not logically restrict the outcome of the response variable. Here, $w = 0$ rules out $y > 0$.
- In the end, we are trying to get flexible models for $D(y|\mathbf{x})$.

3. Truncated Normal Hurdle Model

- Cragg (1971) proposed a natural two-part extension of the type I Tobit model. The conditional independence assumption is assumed to hold, and the binary variable w is assumed to follow a probit model:

$$P(w = 1|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma}).$$

- Further, y^* is assumed to have a *truncated normal distribution* with parameters that vary freely from those in the probit. Can write

$$y^* = \mathbf{x}\boldsymbol{\beta} + u$$

where u given $\mathbf{x}$ has a truncated normal distribution with lower truncation point $-\mathbf{x}\boldsymbol{\beta}$.

- Because $y = y^*$ when $y > 0$, we can write the truncated normal assumption in terms of the density of y given $y > 0$ (and $\mathbf{x}$):

$$f(y|\mathbf{x}, y > 0) = [\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)]^{-1}\phi[(y - \mathbf{x}\boldsymbol{\beta})/\sigma]/\sigma, \quad y > 0,$$

where the term $[\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)]^{-1}$ ensures that the density integrates to unity over $y > 0$.

- The density of y given $\mathbf{x}$ can be written succinctly as

$$f(y|\mathbf{x}) = [1 - \Phi(\mathbf{x}\boldsymbol{\gamma})]^{1[y=0]}\{\Phi(\mathbf{x}\boldsymbol{\gamma})[\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)]^{-1}\phi[(y - \mathbf{x}\boldsymbol{\beta})/\sigma]/\sigma\}^{1[y>0]},$$

where we must multiply $f(y|\mathbf{x}, y > 0)$ by $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})$.

• Called the *truncated normal hurdle (THN) model*. Cragg (1971) directly specified the density.

• Nice feature of the TNH model: it reduces to the type I Tobit model when $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$.

• The log-likelihood function for a random draw i is

$$l_i(\boldsymbol{\theta}) = 1[y_i = 0]\log[1 - \Phi(\mathbf{x}_i\boldsymbol{\gamma})] + 1[y_i > 0]\log[\Phi(\mathbf{x}_i\boldsymbol{\gamma})]$$
$$+ 1[y_i > 0]\{-\log[\Phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma)] + \log\{\phi[(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma]\} - \log(\sigma)\}.$$

Because the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and σ are allowed to freely vary, the MLE for $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}$, is simply the probit estimator from probit of $w_i \equiv 1[y_i > 0]$ on $\mathbf{x}_i$. The MLEs of $\boldsymbol{\beta}$ and σ (or $\boldsymbol{\beta}$ and σ^2) are the MLEs from a truncated normal regression.

• The conditional expectation has the same form as the Type I Tobit because $D(y|\mathbf{x}, y > 0)$ is identical in the two models:

$$E(y|\mathbf{x}, y > 0) = \mathbf{x}\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma).$$

• In particular, the effect of x_j has the same sign as β_j (for continous or discrete changes).

• But now, the relative effect of two continuous variables on the participation probabilities, γ_j/γ_h, can be completely different from β_j/β_h, the ratio of partial effects on $E(y|\mathbf{x}, y > 0)$.

- The unconditional expectation for the Cragg model is

$$E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})[\mathbf{x}\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)].$$

The partial effects no longer have a simple form, but they are not too difficult to compute:

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \gamma_j\phi(\mathbf{x}\boldsymbol{\gamma})[\mathbf{x}\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)] + \Phi(\mathbf{x}\boldsymbol{\gamma})\beta_j\theta(\mathbf{x}\boldsymbol{\beta}/\sigma),$$

where $\theta(z) = 1 - \lambda(z)[z + \lambda(z)]$.

- Note that

$$\log[E(y|x)] = \log[\Phi(x\gamma)] + \log[E(y|\mathbf{x}, y > 0)].$$

• The semi-elasticity with respect to x_j is 100 times

$$\gamma_j \lambda(\mathbf{x}\boldsymbol{\gamma}) + \beta_j \theta(\mathbf{x}\boldsymbol{\beta}/\sigma)/[\mathbf{x}\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]$$

• If $x_j = \log(z_j)$, then the above expression is the elasticity of $\mathrm{E}(y|\mathbf{x})$ with respect to z_j.

• We can insert the MLEs into any of the equations and average across $\mathbf{x}_i$ to obtain an average partial effect, average semi-elastisticity, or average elasticity. As in many nonlinear contexts, the bootstrap is a convienent method for obtaining valid standard errors.

• Can get goodness-of-fit measures as before. For example, the squared correlation between y_i and $\hat{E}(y_i|\mathbf{x}_i) = \Phi(\mathbf{x}_i\hat{\boldsymbol{\gamma}})[\mathbf{x}_i\hat{\boldsymbol{\beta}} + \hat{\sigma}\lambda(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})]$.

4. Lognormal Hurdle Model

- Cragg (1971) also suggested the lognormal distribution conditional on a positive outcome. One way to express y is

$$y = w \cdot y^* = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0]\exp(\mathbf{x}\boldsymbol{\beta} + u),$$

where (u, v) is independent of $\mathbf{x}$ with a bivariate normal distribution; further, u and v are independent.

- y^* has a lognormal distribution because

$$\begin{aligned} y^* &= \exp(\mathbf{x}\boldsymbol{\beta} + u) \\ & u|\mathbf{x} \sim Normal(0, \sigma^2). \end{aligned}$$

Called the lognormal hurdle (LH) model.

- The expected value conditional on $y > 0$ is

$$E(y|\mathbf{x}, y > 0) = E(y^*|\mathbf{x}, w = 1) = E(y^*|\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\beta} + \sigma^2/2).$$

- The semi-elasticity of $E(y|\mathbf{x}, y > 0)$ with respect to x_j is $100\beta_j$. If $x_j = \log(z_j)$, β_j is the elasticity of $E(y|\mathbf{x}, y > 0)$ with respect to z_j.

- The “unconditional” expectation is

$$E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma}) \exp(\mathbf{x}\boldsymbol{\beta} + \sigma^2/2).$$

- The semi-elasticity of E($y|\mathbf{x}$) with respect to x_j is simply (100 times) $\gamma_j \lambda(\mathbf{x}\boldsymbol{\gamma}) + \beta_j$ where $\lambda(\bullet)$ is the inverse Mills ratio. If $x_j = \log(z_j)$, this expression becomes the elasticity of $E(y|\mathbf{x})$ with respect to z_j.

• Estimation of the parameters is particularly straightforward. The density conditional on $\mathbf{x}$ is

$$f(y|\mathbf{x}) = [1 - \Phi(\mathbf{x}\boldsymbol{\gamma})]^{1[y=0]}\{\Phi(\mathbf{x}\boldsymbol{\gamma})\phi[(\log(y) - \mathbf{x}\boldsymbol{\beta})/\sigma]/(\sigma y)\}^{1[y>0]},$$

which leads to the log-likelihood function for a random draw:

$$l_i(\boldsymbol{\theta}) = 1[y_i = 0]\log[1 - \Phi(\mathbf{x}_i\boldsymbol{\gamma})] + 1[y_i > 0]\log[\Phi(\mathbf{x}_i\boldsymbol{\gamma})]$$
$$+ 1[y_i > 0]\{\log(\phi[(\log(y_i) - \mathbf{x}_i\boldsymbol{\beta})/\sigma]) - \log(\sigma) - \log(y_i)\}.$$

• As with the truncated normal hurdle model, estimation of the parameters can proceed in two steps. The first is probit of w_i on $\mathbf{x}_i$ to estimate $\boldsymbol{\gamma}$, and then $\boldsymbol{\beta}$ is estimated using an OLS regression of $\log(y_i)$ on $\mathbf{x}_i$ for observations with $y_i > 0$.

- The usual error variance estimator (or without the degrees-of-freedom adjustment), $\hat{\sigma}^2$, is consistent for σ^2.
- In computing the log likelihood to compare fit across models, must include the terms $\log(y_i)$. In particular, for comparing with the TNH model.
- Can relax the lognormality assumption if we are satisfied with estimates of $P(y > 0|\mathbf{x})$, $E(y|\mathbf{x}, y > 0)$, and $E(y|\mathbf{x})$ are easy to obtain.

• Nevertheless, if we are mainly interested in these three features of $D(y|\mathbf{x})$, we can get by with weaker assumptions. If in $y^* = \exp(\mathbf{x}\boldsymbol{\beta} + u)$ we assume that u is independent of $\mathbf{x}$, can use Duan's (1983) *smearing estimate.*

• Uses $E(y^*|\mathbf{x}) = E[\exp(u)]\exp(\mathbf{x}\boldsymbol{\beta}) \equiv \tau\exp(\mathbf{x}\boldsymbol{\beta})$ where $\tau \equiv E[\exp(u)]$.

• Let $\hat{u}_i$ be OLS residuals from $\log(y_i)$ on $\mathbf{x}_i$ using the $y_i > 0$ data. Let

$$\hat{\tau} = N^{-1}\sum_{i=1}^{N}\exp(\hat{u}_i).$$

Then, $\hat{E}(y|\mathbf{x}, y > 0) = \hat{\tau}\exp(\mathbf{x}\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\log(y_i)$ on $\mathbf{x}_i$ using the $y_i > 0$ subsample.

- More direct approach: just specify

$$E(y|\mathbf{x}, y > 0) = \exp(\mathbf{x}\boldsymbol{\beta}),$$

which contains $y^* = \exp(\mathbf{x}\boldsymbol{\beta} + u),$ with u independent of $\mathbf{x}$, as a special case.

- Use nonlinear least squares or a quasi-MLE in the linear exponential family (such as the Poisson or gamma).

- Given probit estimates of $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})$ and QMLE estimates of $E(y|\mathbf{x}, y > 0) = \exp(\mathbf{x}\boldsymbol{\beta})$, can easily estimate $E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})\exp(\mathbf{x}\boldsymbol{\beta})$ without additional distributional assumptions.

5. Exponential Type II Tobit Model

- Now allow w and y^* to be dependent after conditioning on observed covariates, $\mathbf{x}$. Seems natural – for example, unobserved factors that affect labor force participation can affect amount of hours.
- Can modify the lognormal hurdle model to allow conditional correlation between w and y^*. Call the resulting model the *exponential type II Tobit (ET2T) model*.
- Traditionally, the type II Tobit model has been applied to missing data problems – that is, where we truly have a sample selection issue. Here, we use it as a way to obtain a flexible corner solution model.

- As with the lognormal hurdle model,

$$y = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0]\exp(\mathbf{x}\boldsymbol{\beta} + u)$$

We use the qualifier "exponential" to emphasize that we should have $y^* = \exp(\mathbf{x}\boldsymbol{\beta} + u)$.

- Later we will see why it makes no sense to have $y^* = \mathbf{x}\boldsymbol{\beta} + u$, as is often the case in the study of type II Tobit models of sample selection.
- Because v has variance equal to one, $Cov(u,v) = \rho\sigma$, where ρ is the correlation between u and v and $\sigma^2 = Var(u)$.

- Obtaining the log likelihood in this case is a bit tricky. Let $m^* = \log(y^*)$, so that $D(m^*|\mathbf{x})$ is $Normal(\mathbf{x}\boldsymbol{\beta}, \sigma^2)$. Then $\log(y) = m^*$ when $y > 0$. We still have $P(y = 0|\mathbf{x}) = 1 - \Phi(\mathbf{x}\boldsymbol{\gamma})$.
- To obtain the density of y (conditional on $\mathbf{x}$) over strictly positive values, we find $f(y|\mathbf{x}, y > 0)$ and multiply it by $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})$.
- To find $f(y|\mathbf{x}, y > 0)$, we use the change-of-variables formula $f(y|\mathbf{x}, y > 0) = g(\log(y)|\mathbf{x}, y > 0)/y$, where $g(\bullet|\mathbf{x}, y > 0)$ is the density of m^* conditional on $y > 0$ (and $\mathbf{x}$).

• Use Bayes' rule to write

$g(m^*|\mathbf{x}, w = 1) = P(w = 1|m^*, x)h(m^*|x)/P(w = 1|\mathbf{x})$ where $h(m^*|\mathbf{x})$ is the density of m^* given $\mathbf{x}$. Then,

$P(w = 1|x)g(m^*|x, w = 1) = P(w = 1|m^*, \mathbf{x})h(m^*|\mathbf{x})$.

• Write $w = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0] = 1[\mathbf{x}\boldsymbol{\gamma} + (\rho/\sigma)u + e > 0]$, where $v = (\rho/\sigma)u + e$ and $e|\mathbf{x}, u \sim Normal(0, (1 - \rho^2))$. Because $u = m^* - \mathbf{x}\boldsymbol{\beta}$, we have $P(w = 1|m^*, \mathbf{x}) = \Phi([\mathbf{x}\boldsymbol{\gamma} + (\rho/\sigma)(m^* - \mathbf{x}\boldsymbol{\beta})](1 - \rho^2)^{-1/2})$.

- Further, we have assumed that $h(m^*|\mathbf{x})$ is $Normal(\mathbf{x}\boldsymbol{\beta}, \sigma^2)$. Therefore, the density of y given $\mathbf{x}$ over strictly positive y is

$$f(y|\mathbf{x}) = \Phi([\mathbf{x}\boldsymbol{\gamma} + (\rho/\sigma)(y - \mathbf{x}\boldsymbol{\beta})](1 - \rho^2)^{-1/2}))\phi((\log(y) - \mathbf{x}\boldsymbol{\beta})/\sigma)/(\sigma y).$$

- Combining this expression with the density at $y = 0$ gives the log likelihood as

$$\begin{aligned} l_i(\boldsymbol{\theta}) &= 1[y_i = 0]\log[1 - \Phi(\mathbf{x}_i\boldsymbol{\gamma})] \\ &+ 1[y_i > 0]\{\log[\Phi([\mathbf{x}_i\boldsymbol{\gamma} + (\rho/\sigma)(\log(y_i) - \mathbf{x}_i\boldsymbol{\beta})](1 - \rho^2)^{-1/2}) \\ &\qquad + \log[\phi((\log(y_i) - \mathbf{x}_i\boldsymbol{\beta})/\sigma)] - \log(\sigma) - \log(y_i)\}. \end{aligned}$$

- Many econometrics packages have this estimator programmed, although the emphasis is on sample selection problems. To use Heckman sample selection software, one defines $\log(y_i)$ as the variable where the data are "missing" when $y_i = 0$) When $\rho = 0$, we obtain the log likelihood for the lognormal hurdle model from the previous subsection.

- For a true missing data problem, the last term in the log likelihood, $\log(y_i)$, is not included. That is because in sample selection problems the log-likelihood function is only a partial log likelihood. Inclusion of $\log(y_i)$ does not affect the estimation problem, but it does affect the value of the log-likelihood function, which is needed to compare across different models.)

• The ET2T model contains the conditional lognormal model from the previous subsection. But the ET2T model with unknown ρ can be poorly identified if the set of explanatory variables that appears in $y^* = \exp(\mathbf{x}\boldsymbol{\beta} + u)$ is the same as the variables in $w = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0]$.

• Various ways to see the potential problem. Can show that

$$E[\log(y)|\mathbf{x}, y > 0] = \mathbf{x}\boldsymbol{\beta} + \eta\lambda(\mathbf{x}\boldsymbol{\gamma})$$

where $\lambda(\cdot)$ is the inverse Mills ratio and $\eta = \rho\sigma$.

• We know we can estimate $\boldsymbol{\gamma}$ by probit, so this equation nominally identifies $\boldsymbol{\beta}$ and η. But identification is possible only because $\lambda(\bullet)$ is a nonlinear function, but $\lambda(\bullet)$ is roughly linear over much of its range.

• The formula for $E[\log(y)|\mathbf{x}, y > 0]$ suggests a two-step procedure, usually called *Heckman's method* or *Heckit*. First, $\hat{\boldsymbol{\gamma}}$ from probit of w_i on $\mathbf{x}_i$. Second, $\hat{\boldsymbol{\beta}}$ and $\hat{\eta}$ are obtained from OLS of $\log(y_i)$ on $\mathbf{x}_i$, $\lambda(\mathbf{x}_i\hat{\boldsymbol{\gamma}})$ using only observations with $y_i > 0$.

- The correlation between $\hat{\lambda}_i$ can often be very large, resulting in imprecise estimates of $\boldsymbol{\beta}$ and η.
- Can be shown that the unconditional expectation is

$$E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma} + \eta)\exp(\mathbf{x}\boldsymbol{\beta} + \sigma^2/2),$$

which is exactly of the same form as in the LH model (with $\rho = 0$) except for the presence of $\eta = \rho\sigma$. Because $\mathbf{x}$ always should include a constant, η is not separately identified by $E(y|\mathbf{x})$ (and neither is $\sigma^2/2$).

- If we based identification entirely on $E(y|\mathbf{x})$, there would be no difference between the lognormal hurdle model and the ET2T model when the same set of regressors appears in the participation and amount equations.
- Still, the parameters are technically identified, and so we can always try to estimate the full model with the same vector $\mathbf{x}$ appearing in the participation and amount equations.

• The ET2T model is more convincing when the covariates determining the participation decision strictly contain those affecting the amount decision. Then, the model can be expressed as

$$y = 1(\mathbf{x}\boldsymbol{\gamma} + v \geq 0) \cdot \exp(\mathbf{x}_1\boldsymbol{\beta}_1 + u),$$

where both $\mathbf{x}$ and $\mathbf{x}_1$ contain unity as their first elements but $\mathbf{x}_1$ is a strict subset of $\mathbf{x}$. If we write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, then we are assuming $\boldsymbol{\gamma}_2 \neq \mathbf{0}$.

• Given at least one exclusion restriction, we can see from $E[\log(y)|\mathbf{x}, y > 0] = \mathbf{x}_1\boldsymbol{\beta}_1 + \eta\lambda(\mathbf{x}\boldsymbol{\gamma})$ that $\boldsymbol{\beta}_1$ and η are better identified because $\lambda(\mathbf{x}\boldsymbol{\gamma})$ is not an exact function of $\mathbf{x}_1$.

- Exclusion restrictions can be hard to come by. Need something affecting the fixed cost of participating but not affecting the amount.
- Cannot use y rather than $\log(y)$ in the amount equation. In the TNH model, the truncated normal distribution of u at the value $-\mathbf{x}\boldsymbol{\beta}$ ensures that $y^* = \mathbf{x}\boldsymbol{\beta} + u > 0$.
- If we apply the type II Tobit model directly to y, we must assume (u, v) is bivariate normal and *independent* of $\mathbf{x}$. What we gain is that u and v can be correlated, but this comes at the cost of not specifying a proper density because the T2T model allows negative outcomes on y.

- If we apply the “selection” model to y we would have

$$E(y|\mathbf{x}, y > 0) = \mathbf{x}\boldsymbol{\beta} + \eta\lambda(\mathbf{x}\boldsymbol{\gamma}).$$

- Possible to get negative values for $E(y|\mathbf{x}, y > 0)$, especially when $\rho < 0$. It only makes sense to apply the T2T model to $\log(y)$ in the context of two-part models.

- Example of Two-Part Models: Married Women’s Labor Supply

	(1)	(2)	(3)
Model	Truncated Normal	Lognormal	Exponential
	Hurdle	Hurdle	Type II Tobit
Participation Equation			
nwifeinc	–.012 (.005)	–.012 (.005)	–.0097 (.0043)
educ	.131 (.025)	.131 (.025)	.120 (.022)
exper	.123 (.019)	.123 (.019)	.083 (.017)
$exper^2$	–.0019 (.0006)	–.0019 (.0006)	–.0013 (.0005)
age	–.088 (.015)	–.088 (.015)	–.033 (.008)
*kidslt*6	–.868 (.119)	–.868 (.119)	–.504 (.107)
*kidsge*6	.036 (.043)	.036 (.043)	.070 (.039)
constant	.270 (.509)	.270 (.509)	–.367 (.448)

	(1)	(2)	(3)
Model	Truncated Normal	Lognormal	Exponential
	Hurdle	Hurdle	Type II Tobit
Amount Equation	*hours*	log(*hours*)	log(*hours*)
nwifeinc	.153 (5.164)	–.0020 (.0044)	.0067 (.0050)
educ	–29.85 (22.84)	–.039 (.020)	–.119 (.024)
exper	72.62 (21.24)	.073 (.018)	–.033 (.020)
$exper^2$	–.944 (.609)	–.0012 (.0005)	.0006 (.0006)
age	–27.44 (8.29)	–.024 (.007)	.014 (.008)
*kidslt*6	–484.91 (153.79)	–.585 (.119)	.208 (.134)
*kidsge*6	–102.66 (43.54)	–.069 (.037)	–.092 (.043)
constant	2,123.5 (483.3)	7.90 (.43)	8.67 (.50)

	(1)	(2)	(3)
Model	Truncated Normal	Lognormal	Exponential
	Hurdle	Hurdle	Type II Tobit
$\hat{\sigma}$	850.77 (43.80)	.884 (.030)	1.209 (.051)
$\hat{\rho}$	—	—	−.972 (.010)
Log Likelihood	−3,791.95	−3,894.93	−3,877.88
Number of Women	753	753	753

```
. * use mroz

. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(7)      =       227.
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.30219                       Pseudo R2       =     0.2206

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+----------------------------------------------------------------
    nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
        educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
     expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
         age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
     kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
     kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
       _cons |   .2700768    .508593     0.53   0.595    -.7267473    1.266901
------------------------------------------------------------------------------
```

```
. truncreg hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)
(note: 325 obs. truncated)

Truncated regression
Limit:   lower =          0                               Number of obs =    428
         upper =       +inf                               Wald chi2(7)  =  59.
Log likelihood = -3390.6476                               Prob > chi2   = 0.0000

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+----------------------------------------------------------------
    nwifeinc |   .1534399   5.164279     0.03   0.976    -9.968361    10.27524
        educ |  -29.85254   22.83935    -1.31   0.191    -74.61684    14.91176
       exper |   72.62273   21.23628     3.42   0.001     31.00039    114.2451
     expersq |  -.9439967   .6090283    -1.55   0.121     -2.13767    .2496769
         age |  -27.44381   8.293458    -3.31   0.001    -43.69869   -11.18893
     kidslt6 |  -484.7109   153.7881    -3.15   0.002      -786.13   -183.2918
     kidsge6 |  -102.6574   43.54347    -2.36   0.018    -188.0011   -17.31379
       _cons |   2123.516   483.2649     4.39   0.000     1176.334    3070.697
-------------+----------------------------------------------------------------
      /sigma |    850.766   43.80097    19.42   0.000     764.9177    936.6143
------------------------------------------------------------------------------

. * log likelihood for Cragg truncated normal hurdle model

. di -3390.6476 - 401.30219
-3791.9498


. * A trick to get the log likelihood for the lognormal hurdle model:
```

```
. tobit lhours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)

Tobit regression                                  Number of obs   =        428
                                                  LR chi2(7)      =      77.
                                                  Prob > chi2     =     0.0000
Log likelihood = -554.56647                       Pseudo R2       =     0.0653

------------------------------------------------------------------------------
      lhours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
    nwifeinc |  -.0019676   .0044019    -0.45   0.655      -.01062    .0066848
        educ |  -.0385626     .02002    -1.93   0.055    -.0779142    .0007891
       exper |    .073237   .0177323     4.13   0.000     .0383821    .1080919
     expersq |   -.001233   .0005328    -2.31   0.021    -.0022803   -.0001858
         age |  -.0236706   .0071799    -3.30   0.001    -.0377836   -.0095576
     kidslt6 |   -.585202   .1174928    -4.98   0.000    -.8161477   -.3542563
     kidsge6 |  -.0694175   .0369849    -1.88   0.061    -.1421156    .0032806
       _cons |   7.896267   .4220778    18.71   0.000     7.066625     8.72591
-------------+----------------------------------------------------------------
      /sigma |    .884067   .0302167                      .8246725    .9434614
------------------------------------------------------------------------------
  Obs. summary:          0  left-censored observations
                       428     uncensored observations
                         0 right-censored observations
```

```
. * log likelihood for lognormal hurdle:

. sum lhours

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      lhours |       428     6.86696    .9689285   2.484907   8.507143


. di -401.30219 - 554.56647  - 428*6.86696
-3894.9275

. * Now get the llf for each nonzero observation to compute the Vuong
. * test for the truncated normal versus lognormal.

. predict xb1
(option xb assumed; fitted values)

. gen llf1 = log(normalden((lhours - xb1)/.88407)) - log(.88407) - lhours
(325 missing values generated)
```

```
. truncreg hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)
(note: 325 obs. truncated)

Truncated regression
Limit:   lower =          0                            Number of obs =    428
         upper =       +inf                            Wald chi2(7)  =  59.
Log likelihood = -3390.6476                            Prob > chi2   = 0.0000

---------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+-------------------------------------------------------------
    nwifeinc |   .1534399   5.164279     0.03   0.976    -9.968361    10.27524
        educ |  -29.85254   22.83935    -1.31   0.191    -74.61684    14.91176
       exper |   72.62273   21.23628     3.42   0.001     31.00039    114.2451
     expersq |  -.9439967   .6090283    -1.55   0.121     -2.13767    .2496769
         age |  -27.44381   8.293458    -3.31   0.001    -43.69869   -11.18893
     kidslt6 |  -484.7109   153.7881    -3.15   0.002      -786.13   -183.2918
     kidsge6 |  -102.6574   43.54347    -2.36   0.018    -188.0011   -17.31379
       _cons |   2123.516   483.2649     4.39   0.000     1176.334    3070.697
-------------+-------------------------------------------------------------
      /sigma |    850.766   43.80097    19.42   0.000     764.9177    936.6143
---------------------------------------------------------------------------
```

```
. predict xb2, xb

. gen u2 = hours - xb2

. gen llf2 = log(normalden(u2/ 850.766 )) - log( 850.766 )
     - log(normal(xb2/ 850.766))

. replace llf2 = . if hours == 0
(325 real changes made, 325 to missing)

. gen diff = llf2 - llf1
(325 missing values generated)
```

```
. reg diff

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  0,   427) =    0.
       Model |           0     0           .           Prob > F      =
    Residual |  203.969251   427  .477679746           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |  203.969251   427  .477679746           Root MSE      =  .69114

------------------------------------------------------------------------------
        diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
       _cons |   .2406023   .0334077     7.20   0.000     .1749383    .3062663
------------------------------------------------------------------------------

. * The Vuong test strongly rejects the lognormal in favor of the truncated
. * in terms of fit.
```

```
. heckman lhours nwifeinc educ exper expersq age kidslt6 kidsge6,
    select(inlf = nwifeinc educ exper expersq age kidslt6 kidsge6)

Iteration 0:   log likelihood = -956.85771
Iteration 1:   log likelihood = -952.20425
Iteration 2:   log likelihood = -940.24444
Iteration 3:   log likelihood = -938.83566
Iteration 4:   log likelihood = -938.82081
Iteration 5:   log likelihood =  -938.8208

Heckman selection model                         Number of obs      =       753
(regression model with sample selection)        Censored obs       =       325
                                                Uncensored obs     =       428

                                                Wald chi2(7)       =     35.
Log likelihood = -938.8208                      Prob > chi2        =    0.0000

----------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+--------------------------------------------------------------
lhours       |
    nwifeinc |   .0066597   .0050147     1.33   0.184    -.0031689    .0164882
        educ |  -.1193085   .0242235    -4.93   0.000    -.1667858   -.0718313
       exper |  -.0334099   .0204429    -1.63   0.102    -.0734773    .0066574
     expersq |   .0006032   .0006178     0.98   0.329    -.0006077    .0018141
         age |   .0142754   .0084906     1.68   0.093    -.0023659    .0309167
     kidslt6 |   .2080079   .1338148     1.55   0.120    -.0542643    .4702801
     kidsge6 |  -.0920299   .0433138    -2.12   0.034    -.1769235   -.0071364
       _cons |   8.670736    .498793    17.38   0.000      7.69312    9.648352
-------------+--------------------------------------------------------------
```

```
inlf         |
    nwifeinc |  -.0096823   .0043273    -2.24   0.025    -.0181637    -.001201
        educ |    .119528   .0217542     5.49   0.000     .0768906    .1621654
       exper |   .0826696   .0170277     4.86   0.000      .049296    .1160433
     expersq |  -.0012896   .0005369    -2.40   0.016     -.002342   -.0002372
         age |  -.0330806   .0075921    -4.36   0.000    -.0479609   -.0182003
     kidslt6 |  -.5040406   .1074788    -4.69   0.000    -.7146951    -.293386
     kidsge6 |   .0698201   .0387332     1.80   0.071    -.0060955    .1457357
       _cons |  -.3656166   .4476569    -0.82   0.414    -1.243008    .5117748
-------------+----------------------------------------------------------------
     /athrho |  -2.131542    .174212   -12.24   0.000    -2.472991   -1.790093
    /lnsigma |   .1895611   .0419657     4.52   0.000     .1073099    .2718123
-------------+----------------------------------------------------------------
         rho |  -.9722333   .0095403                     -.9858766   -.9457704
       sigma |   1.208719   .0507247                      1.113279    1.312341
      lambda |  -1.175157   .0560391                     -1.284991   -1.065322
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    34.10   Prob > chi2 = 0.0000
------------------------------------------------------------------------------

. sum lhours

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      lhours |       428     6.86696    .9689285   2.484907   8.507143

. * log likelihood for the "selection" model:

. di -938.8208 - 428*6.86696
-3877.8797
```