TWO-PART AND HURDLE MODELS

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

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

1. INTRODUCTION

- We consider the case with a corner at zero and a continuous distribution for strictly positive values.
- 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 s be a binary variable that determines whether y is zero or strictly positive. Let w^* be a nonnegative, continuous random variable. Assume y is generated as

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

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

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

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

• This assumption typically underlies *two-part* or *hurdle* models.

• One implication is that the expected value of y conditional on $\mathbf{x}$ and s is easy to obtain:

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

- Sufficient is conditional mean independence,

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

- When $s = 1$, we can write

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

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

- The so-called "unconditional" expectation is

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

- 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 (s in the current notation) does not logically restrict the outcome of the response variable. Here, $s = 0$ 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 s is assumed to follow a probit model:

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

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

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

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

• Because $y = w^*$ 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 (TNH) 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

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

- 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 $s_i \equiv 1[y_i > 0]$ on $\mathbf{x}_i$. The MLEs of $\boldsymbol{\beta}$ and σ (or $\boldsymbol{\beta}$ and σ^2) are the MLEs from what is called 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

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

• 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

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

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

- w^* has a lognormal distribution because

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

Called the lognormal hurdle (LH) model.

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

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

- 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:

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

• As with the truncated normal hurdle model, estimation of the parameters can proceed in two steps. The first is probit of s_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.

• The second-part models can be formally compared using Vuong's (1988, *Econometrica*) *model selection statistic*.

• Vuong's approach applies to models that are nonnested. The null hypothesis is that, in the population, each model fits the data equally well, and therefore both models are necessarily misspecified.

• Let $\boldsymbol{\theta}_1^*$ be the plim of the quasi-MLE from the first model and $\boldsymbol{\theta}_2^*$ the plim of the QMLE from the second model. Then the null is

$$H_0 : E[\ell_{i1}(\boldsymbol{\theta}_1^*)] = E[\ell_{i2}(\boldsymbol{\theta}_2^*)]$$

• Of course, if model 1 is correctly specified, $E[\ell_{i1}(\boldsymbol{\theta}_1^*)] > E[\ell_{i2}(\boldsymbol{\theta}_2^*)]$ (and we usually denote $\boldsymbol{\theta}_1^*$ as $\boldsymbol{\theta}_{o1}$).

• Importantly, the Vuong test allows us to only reject one model against another; we cannot conclude we have the correct model.

- The statistic is based on the asymptotic distribution of

$$N^{-1/2}(\mathcal{L}_1 - \mathcal{L}_2) = N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\hat{\boldsymbol{\theta}}_1) - \ell_{i2}(\hat{\boldsymbol{\theta}}_2)]$$

Assuming standard regularity conditions, it can be shown via a standard mean-value expansion that

$$N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\hat{\boldsymbol{\theta}}_1) - \ell_{i2}(\hat{\boldsymbol{\theta}}_2)] = N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\boldsymbol{\theta}_1^*) - \ell_{i2}(\boldsymbol{\theta}_2^*)] + o_p(1).$$

(See Problem 13.13.)

- This representation is useful when the models are nonnested because then $P[\ell_{i1}(\boldsymbol{\theta}_1^*)] \neq \ell_{i2}(\boldsymbol{\theta}_2^*)] > 0$, and so $\ell_{i1}(\boldsymbol{\theta}_1^*) - \ell_{i2}(\boldsymbol{\theta}_2^*)$ is not identically equal to zero. Under H_0, it does have a zero mean.
- We can apply the CLT directly:

$$N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\boldsymbol{\theta}_1^*) - \ell_{i2}(\boldsymbol{\theta}_2^*)] \stackrel{d}{\rightarrow} Normal(0,\eta^2)$$

$$\eta^2 \equiv Var(d_i^*)$$

where $d_i^* \equiv \ell_{i1}(\boldsymbol{\theta}_1^*) - \ell_{i2}(\boldsymbol{\theta}_2^*)$.

- One version of the test statistic is

$$VMS = \frac{N^{-1/2}\sum_{i=1}^{N}[\ell_{i1}(\hat{\boldsymbol{\theta}}_1) - \ell_{i2}(\hat{\boldsymbol{\theta}}_2)]}{\left\{N^{-1}\sum_{i=1}^{N}[\ell_{i1}(\hat{\boldsymbol{\theta}}_1) - \ell_{i2}(\hat{\boldsymbol{\theta}}_2)]^2\right\}^{1/2}} \xrightarrow{d} Normal(0,1)$$

- An easier calculation is to define, for each i

$$\hat{d}_i = \ell_{i1}(\hat{\boldsymbol{\theta}}_1) - \ell_{i2}(\hat{\boldsymbol{\theta}}_2),$$

the difference in estimated log likelihoods for each i. Then, just do a test that the mean is zero: under the null, the estimation of $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$ has no effect asymptotically. We can use regress $\hat{d}_i$ on 1 and use a standard t test.

- For the current application, we only use the nonlimit observations, that is, $y_i > 0$.

```
. use mroz

. * Compute Vuong test for truncated normal versus lognormal. Because the
. * probit parts are the same, it does not play a role in the test. It does
. * for computing partial effects on the unconditional mean and for
. * comparing the log-likelihoods with other models.

. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Probit regression                                 Number of obs   =        753
                                                  LR chi2(7)      =     227.14
                                                  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
------------------------------------------------------------------------------
```

```
. * Do LH model first:

. reg lhours nwifeinc educ exper expersq age kidslt6 kidsge6

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  7,   420) =   11.90
       Model |  66.3633428     7  9.48047755           Prob > F      =  0.0000
    Residual |  334.513835   420  .796461511           R-squared     =  0.1655
-------------+------------------------------           Adj R-squared =  0.1516
       Total |  400.877178   427   .93882243           Root MSE      =  .89245

------------------------------------------------------------------------------
      lhours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0019676   .0044436    -0.44   0.658    -.0107021    .0067668
        educ |  -.0385626   .0202098    -1.91   0.057    -.0782876    .0011624
       exper |    .073237   .0179004     4.09   0.000     .0380514    .1084225
     expersq |   -.001233   .0005378    -2.29   0.022    -.0022902   -.0001759
         age |  -.0236706    .007248    -3.27   0.001    -.0379175   -.0094237
     kidslt6 |   -.585202   .1186066    -4.93   0.000    -.8183386   -.3520654
     kidsge6 |  -.0694175   .0373355    -1.86   0.064    -.1428053    .0039703
       _cons |   7.896267   .4260789    18.53   0.000     7.058755     8.73378
------------------------------------------------------------------------------
```

```
. predict xb1
(option xb assumed; fitted values)

. predict u1, resid
(325 missing values generated)

. di sqrt(421/428)*.89245
.88512184

. * It is important to make sure we compute the LLF for the lognormal
. * distribution, which means subtracting log(hours):

. gen llf1 = log(normalden(u1/.88512184)) - log(.88512184) - lhours
(325 missing values generated)

. sum llf1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        llf1 |       428   -8.162678    .8146383  -12.79851   -6.26466

. di 428*-8.162678
-3493.6262
```

```
. * So the LH log likelihood for the positive part is -3,493.63

. * Now for the truncated normal:

. 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)  =  59.05
Log likelihood = -3390.6476                             Prob > chi2   = 0.0000

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
    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        |
       _cons |    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 ~inlf
(325 real changes made, 325 to missing)

. sum llf2

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        llf2 |       428   -7.922074    .7561236  -15.55169  -6.853047

. di 428*-7.922074
-3390.6477
```

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

. reg diff

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  0,   427) =    0.00
       Model |           0     0           .           Prob > F      =       .
    Residual |  203.606866   427  .476831069           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |  203.606866   427  .476831069           Root MSE      =  .69053

------------------------------------------------------------------------------
        diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2406037    .033378     7.21   0.000     .1749981    .3062094
------------------------------------------------------------------------------

. * The truncated normal fits substantially better, and we can reject the
. * lognormal very strongly.
```

```
. gen yh2 = xb2 + 850.766*(normden(xb2/ 850.766)/norm(xb2/ 850.766))

. replace yh2 = . if hours == 0
(325 real changes made, 325 to missing)

. gen yh1 = exp(xb1 + ((.88512184)^2)/2)

. replace yh1 = . if hour == 0
(325 real changes made, 325 to missing)

. corr hours yh1
(obs=428)

             |    hours      yh1
-------------+------------------
       hours |   1.0000
         yh1 |   0.3579   1.0000

. di .3579^2
.12809241
```

```
. corr hours yh2
(obs=428)

             |    hours      yh2
-------------+------------------
       hours |   1.0000
         yh2 |   0.3723   1.0000


. di .3723^2
.13860729

. * So the truncated normal fits the conditional mean,
. * E(hours|x,hours > 0), somewhat better, too.
. * What we have not verified is whether the estimated partial effects on
. * E(hours|x,hours > 0) are much different across the models.
```

• If we are mainly interested in $P(y > 0|\mathbf{x})$, $E(y|\mathbf{x}, y > 0)$, and $E(y|\mathbf{x})$, then we can relax the lognormality assumption in the TNH.

• If in $w^* = \exp(\mathbf{x}\boldsymbol{\beta} + u)$ we assume that u is independent of $\mathbf{x}$, can use Duan's (1983) *smearing estimate.*

• Uses $E(w^*|\mathbf{x}) = E[\exp(u)]\exp(\mathbf{x}\boldsymbol{\beta}) \equiv \tau\exp(\mathbf{x}\boldsymbol{\beta})$ 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. Suppose the y_i observations are the first N_1 observations.

- Let

$$\hat{\tau} = N_1^{-1} \sum_{i=1}^{N_1} \exp(\hat{u}_i).$$

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.

• A more direct approach is to just specify

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

which contains $w^* = \exp(\mathbf{x}\boldsymbol{\beta} + u),$ 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, which we will cover in EC 821B).

• Given probit estimates of $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})$ and NLS or 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. Computation of semi-elasticities and elasticities follows along the same lines as under the homoskedastic lognormality assumption.

5. EXPONENTIAL TYPE II TOBIT MODEL

- Now allow s and w^* 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 s and w^*. 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 the latent variable is $w^* = \exp(\mathbf{x}\boldsymbol{\beta} + u)$.

- Later we will see why it makes no sense to have $w^* = \mathbf{x}\boldsymbol{\beta} + u$, 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 $m^* = \log(w^*)$, 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 $g(\cdot|\mathbf{x}, y > 0)$ is the density of m^* conditional on $y > 0$ (and $\mathbf{x}$).

• Use Bayes' rule to write

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

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

• Write $s = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0] = 1[\mathbf{x}\boldsymbol{\gamma} + (\rho/\sigma)u + e > 0]$, 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

$$P(s = 1|m^*, \mathbf{x}) = \Phi([\mathbf{x}\boldsymbol{\gamma} + (\rho/\sigma)(m^* - \mathbf{x}\boldsymbol{\beta})](1 - \rho^2)^{-1/2}).$$

- 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

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

• 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, and one must define $\log(y_i)$ 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.

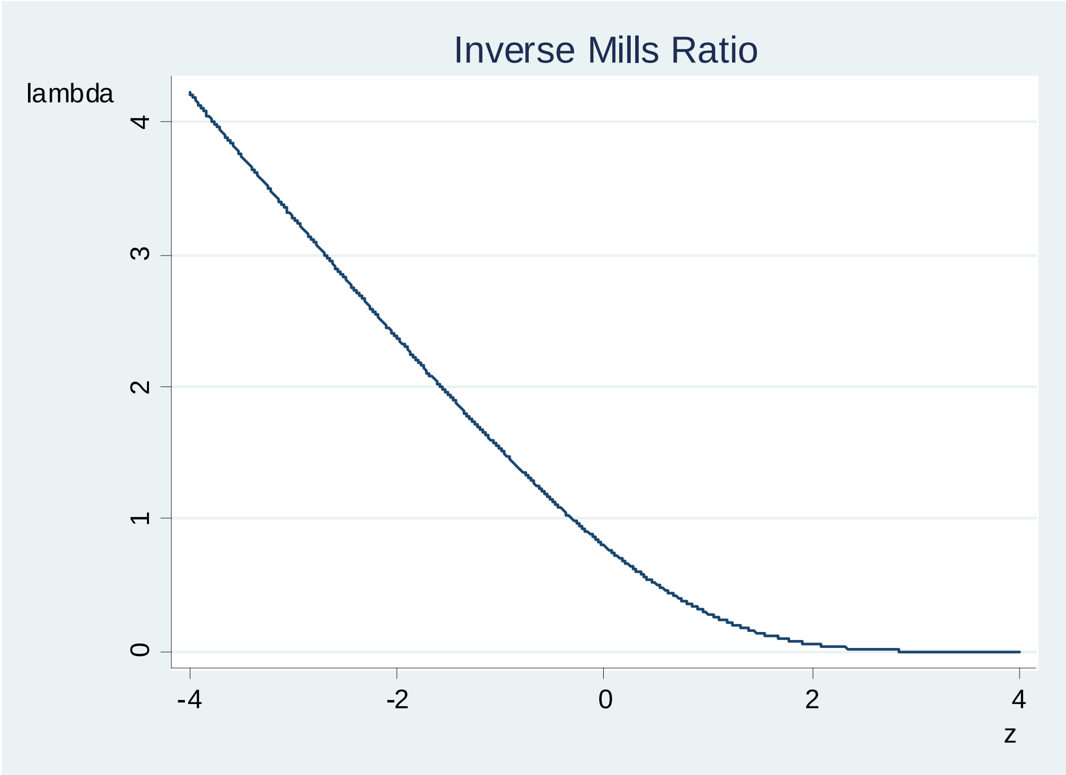
- For a true missing data problem, the last term in the log likelihood, $\log(y_i)$, is not included. That is because in sample selection problems the log-likelihood function is only a partial log likelihood. Inclusion of $\log(y_i)$ does not affect the estimation problem, but it does affect the value 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 $w^* = \exp(\mathbf{x}\boldsymbol{\beta} + u)$ is the same as the variables in $s = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0]$.

- Various ways to see the potential problem. First, 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 consistently estimate $\boldsymbol{\gamma}$ by probit, so this equation nominally identifies $\boldsymbol{\beta}$ and η. But identification is possible only because $\lambda(\cdot)$ is a nonlinear function.

- The identification is tenuous because $\lambda(\cdot)$ is roughly linear over much of its range.


Inverse Mills Ratio
lambda
0
1
2
3
4
-4
-2
0
2
4
z

• The expression for $E[\log(y)|\mathbf{x}, y > 0]$ suggests a two-step procedure, usually called *Heckman's method* or *Heckit*. (Usually used for nonrandom sampling.) (1) Obtain $\hat{\boldsymbol{\gamma}}$ from probit of s_i on $\mathbf{x}_i$. (2) Obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\eta}$ 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.

• In fact, it can be shown that if we replace the probit model for s with a linear probability model then identification of $\boldsymbol{\beta}$ and η is lost. Then

$$s = \mathbf{x}\boldsymbol{\gamma} + v$$

and a natural assumption is $E(u|\mathbf{x}, v) = E(u|v) = \eta v$. The Heckman equation becomes

$$\begin{aligned} E[\log(y)|\mathbf{x}, s = 1] &= \mathbf{x}\boldsymbol{\beta} + \eta v = \mathbf{x}\boldsymbol{\beta} + \eta(1 - \mathbf{x}\boldsymbol{\gamma}) \\ &= \mathbf{x}\boldsymbol{\beta} + \eta - \eta(\mathbf{x}\boldsymbol{\gamma}) \end{aligned}$$

which shows that η and $\boldsymbol{\beta}$ are not identified because $\mathbf{x}$ contains an intercept and $(\mathbf{x}\boldsymbol{\gamma})$ is perfectly collinear with $\mathbf{x}$.

• 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.

- Technically, the parameters are identified, and so we can try to estimate the full model with the same vector **x** appearing in the participation and amount equations. In practice it usually does not work very well. Like other instances of achieving identification off of nonlinearities, it is viewed with skepticism

• Partial effects can be hard to even sign. For the conditional expectation of $\log(y)$,

$$\frac{\partial E[\log(y)|\mathbf{x}, y > 0]}{\partial x_j} = \beta_j + \eta\lambda^{(1)}(\mathbf{x}\boldsymbol{\gamma})\gamma_j$$

where $\lambda^{(1)}(\cdot) < 0$ is the first derivative of the IMR. The sign of η is the same as $\rho = Corr(u, v)$.

• The partial effects on the unconditional expectation of y are

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

which is easy to sign if β_j and γ_j have the same sign, but not otherwise.

- The semi-elasticity is

$$\frac{\partial \log E(y|\mathbf{x})}{\partial x_j} = \gamma_j \lambda(\mathbf{x}\boldsymbol{\gamma} + \eta) + \beta_j$$

which is positive if $\gamma_j,\ \beta_j > 0$ and negative if $\gamma_j,\ \beta_j < 0$. Otherwise, the sign can depend on $\mathbf{x}$.

```
. gen lhours = log(hours)
(325 missing values generated)

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

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

                                                Wald chi2(7)       =     35.50
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
-------------+----------------------------------------------------------------
```

```
select       |
    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

. di -938.8208 - 428*( 6.86696)
-3877.8797

. * This value of the LLF is below the truncated normal hurdle model, which is
. * -3,791.95. Of course, it is above that for the lognormal hurdle model
. * because the ET2T model nests the LNH model (-3,894.93).
```

- The ET2T model is more convincing when the covariates determining the amount are a strict subset of those affecting participation. 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 $w^* = \mathbf{x}\boldsymbol{\beta} + u > 0$.
- 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.