Two-Part Models and Selection Models for Corner Solution Outcomes

2017 FARS Midyear Meeting
Charlotte, NC
January 25, 2017

Jeff Wooldridge
Department of Economics
Michigan State University

1. Preliminaries
2. Review of the Tobit Model
3. Motivation for Two-Part Models
4. A General Formulation
5. Truncated Normal Hurdle Model
6. Lognormal Hurdle Model
7. Applications of Hurdle Models
8. Exponential Type II Tobit Model

1. Preliminaries

- Important to know whether a problem is properly studied in a two-part (hurdle) framework or a self-selection framework.
- If it makes sense to use a Tobit model for y then there is no self-selection problem.
- The practical issue is whether a Tobit model is sufficient. Is a two-part model is warranted?

- **Example**: *y* annual charitable contributions. This is a *corner solution* response, where zero represents a true outcome.
- If a Tobit model is not sufficient, a two-part model might be needed.
- But modeling *y* is not properly viewed as a self-selection problem.

- **Accounting Example**: *y* is the dollar value of equity incentive contracts.
- There is only one outcome, and it might be zero.
- Core and Guay (1999): "We assume that the firm simultaneously chooses to make a grant of equity incentives, and the magnitude of the incentive grant conditional on making a grant."

- In order to be in a self-selection framework, one must be able to define counterfactual outcomes in the two states.
- **Example**: Let w denote participation in a job training program, where some workers participate and others do not.
- y_0 is earnings in the absense of job training, y_1 is earnings with job training.
- For each person, i, we only observe

$$y_i = (1 - w_i)y_{i0} + w_i y_{i1}$$

along with w_i.

- **Accounting Example**: w is one if a German firm adopts international disclosure rules, zero if not.
- y_0 is the cost of capital if the firm does not; y_1 is the cost of capital if the firm does.
- These counterfactuals make sense. (Leuz and Verrecchia, Journal of Accounting Research, 2000).
- Another Example: w is one if a hospital is for profit, zero if not. y is a cost measure. Again, y_0 and y_1 are conceptually well defined.

- **Charitable Contributions Example**: It makes no sense to ask: “How much would a family contribute to charity if it does not contribute to charity?”
- We may want to use a two-part model for charitable contributions, but that is much different from having two potential outcomes.
- Confusion arises because a common two-part model is called a “selection” model.
- See FARS 2016 lecture for IV and control function approaches to self-selection.

2. Review of the Tobit Model

- $y \geq 0$ is a random variable continuous except at zero:

$$P(y = 0) > 0$$

- The most we can know about y is its conditional distribution, $D(y|x)$.

- Often reported in applications are partial effects on

(i) $P(y > 0|\mathbf{x})$

(ii) $E(y|\mathbf{x}, y > 0)$ (the "conditional" expectation)

(iii) $E(y|\mathbf{x})$ (the "unconditional" expectation)

- Not crazy to start with a linear approximation:

$$E(y|\mathbf{x}) \approx \mathbf{x}\boldsymbol{\gamma} = \gamma_1 + \gamma_2 x_2 + \cdots + \gamma_K x_K.$$

- Note: This is a *linear model* that we *estimate* by OLS.
- We do not use the phrase “I estimated an OLS model.”
- The $\hat{\gamma}_j$ might be reasonable approximations to the average partial effects.

- Tobit can give a better functional form for partial effects over a broad range of x_j values.
- y follows a Type I Tobit Model if

$$y = \max(0, \mathbf{x}\boldsymbol{\beta} + u)$$
$$u|x \sim \text{Normal}(0, \sigma^2).$$

- There is no censoring problem! A zero is a zero!

• Quantities of Interest for Tobit:

(i) $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$. If x_j is continuous, we can take a derivative:

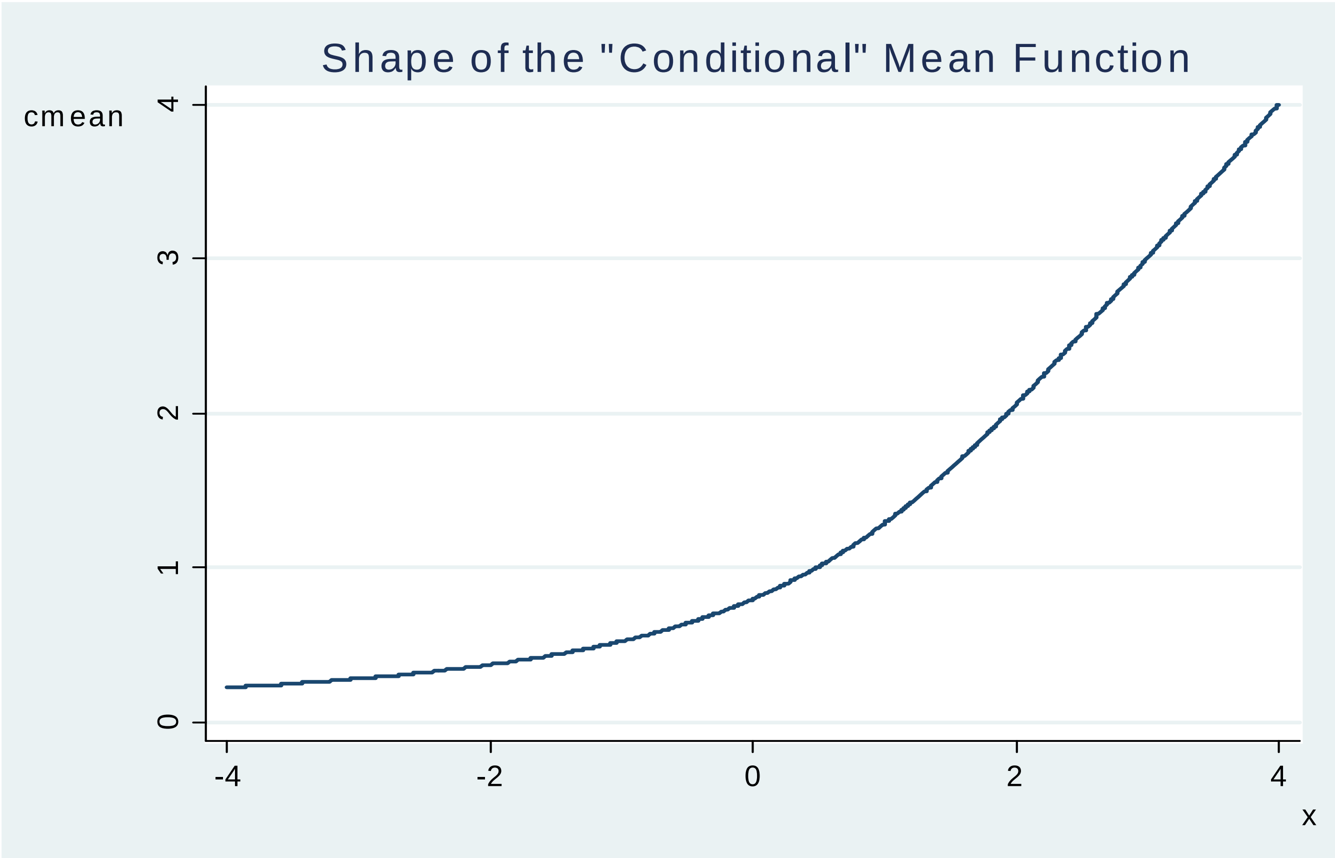
$$\frac{\partial P(y > 0|x)}{\partial x_j} = (\beta_j/\sigma)\phi(\mathbf{x}\boldsymbol{\beta}/\sigma),$$

so the partial effect depends on $\boldsymbol{\beta}$, σ, and all of the elements of x.

(ii) The "conditional" expectation (conditional on $y > 0$):

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

where $\lambda(z) = \phi(z)/\Phi(z)$ is the "inverse Mills ratio" (IMR).

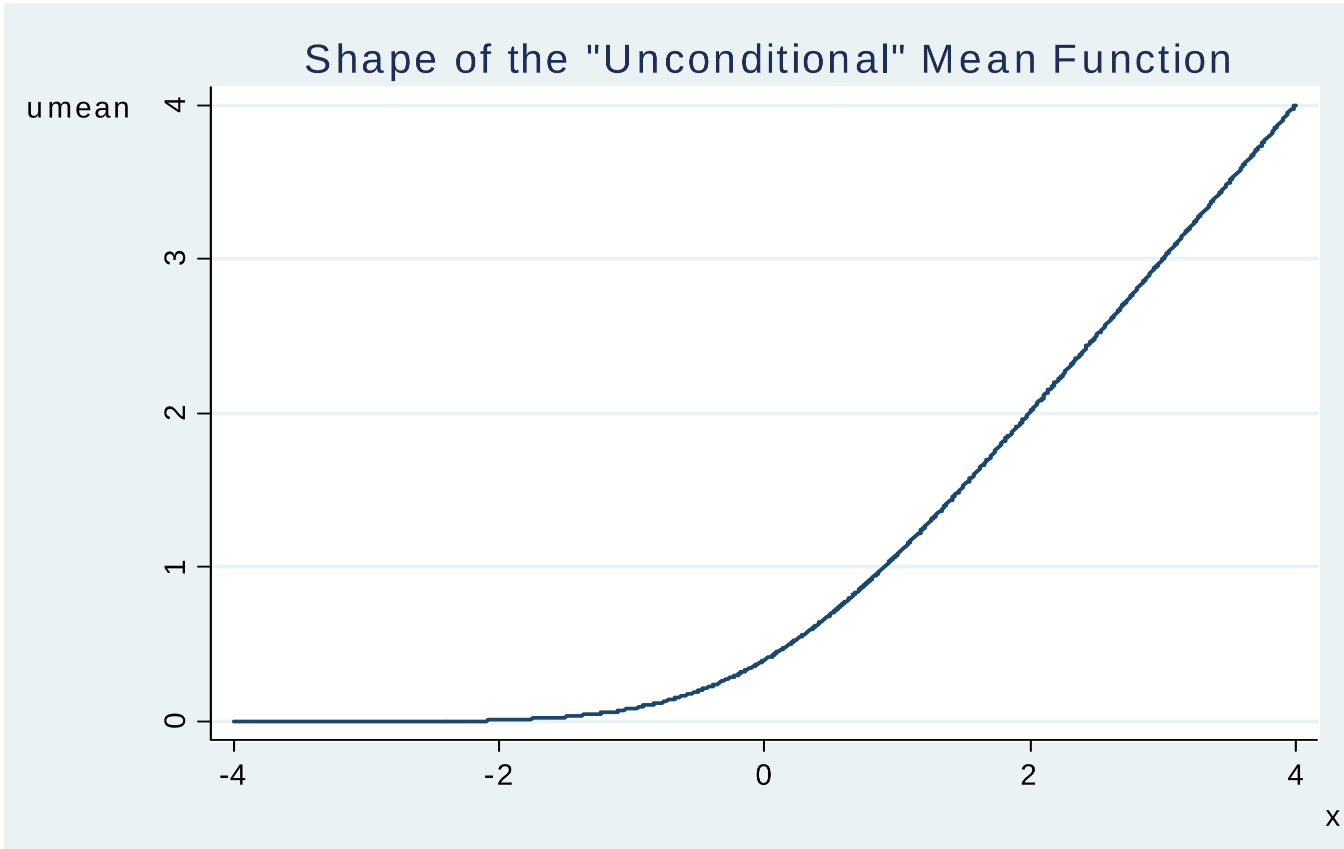

Shape of the "Conditional" Mean Function
cmean
0
1
2
3
4
-4
-2
0
2
4
x

(iii) The "unconditional" expectation:

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

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \beta_j\Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$$

• Note: σ is not "ancillary" for estimating partial effects on for means of y.


Shape of the "Unconditional" Mean Function
umean
4
3
2
1
0
-4
-2
0
2
4
x

• We can evaluate the partial effects at interesting values of x (means, medians, quantiles) but often it is useful to have a single measure, the average partial effect (APE):

$$\widehat{APE}_j = \left[N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}_i \hat{\boldsymbol{\beta}}/\hat{\sigma}) \right] \hat{\beta}_j.$$

• Can compare$\widehat{APE}_j$) with the corresponding OLS coefficient, $\hat{\gamma}_j$.

• Can embellish Tobit in several ways, but it is inherently a “one-part” model.

```
. use charity

. des gift lgift resplast weekslast mailsyear propresp avggift lavggift

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
gift            int    %9.0g                  amount of gift, Dutch guilders
lgift           float  %9.0g                  log(gift); missing if gift = 0
resplast        byte   %9.0g                  =1 if responded to most recent mailing
weekslast       float  %9.0g                  number of weeks since last response
mailsyear       float  %9.0g                  number of mailings per year
propresp        float  %9.0g                  response rate to mailings
avggift         float  %9.0g                  average of past gifts
lavggift        float  %9.0g                  log(avggift)
```

```
. sum gift lgift resplast weekslast mailsyear propresp avggift lavggift

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        gift |      4,268     7.44447    15.06256          0        250
       lgift |      1,707    2.652349    .6998901   .6931472   5.521461
    resplast |      4,268    .3348172    .4719818          0          1
   weekslast |      4,268     59.0482    44.32374   13.14286        195
   mailsyear |      4,268    2.049555      .66758        .25        3.5
-------------+---------------------------------------------------------
    propresp |      4,268    .4843592    .2533932    .090909          1
     avggift |      4,268    18.24284    78.70286          1       5005
    lavggift |      4,268    2.589046    .6699614          0   8.518192
```

```
. reg gift resplast weekslast mailsyear propresp lavggift, robust

Linear regression                               Number of obs     =      4,268
                                                F(5, 4262)        =     118.81
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2335
                                                Root MSE          =     13.195

------------------------------------------------------------------------------
             |               Robust
        gift |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    resplast |   .8323287   .6714167     1.24   0.215    -.4839976    2.148655
   weekslast |   -.017745   .0056432    -3.14   0.002    -.0288086   -.0066813
   mailsyear |   .3447685   .3962544     0.87   0.384    -.4320965    1.121633
    propresp |   13.20904   1.196681    11.04   0.000     10.86292    15.55516
    lavggift |    8.87639   .7987143    11.11   0.000     7.310494    10.44229
       _cons |  -21.87232   1.918434   -11.40   0.000    -25.63345   -18.11119
------------------------------------------------------------------------------
```

```
. tobit gift i.resplast weekslast mailsyear propresp lavggift, ll(0)

Tobit regression                                Number of obs     =      4,268
                                                LR chi2(5)        =    1110.76
                                                Prob > chi2       =     0.0000
Log likelihood = -9115.3322                     Pseudo R2         =     0.0574

------------------------------------------------------------------------------
        gift |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  1.resplast |   .6063508   1.223619     0.50   0.620     -1.79258    3.005282
   weekslast |  -.1277379   .0160405    -7.96   0.000    -.1591857   -.0962901
   mailsyear |    1.54373   .6828978     2.26   0.024     .2048943    2.882565
    propresp |    34.6707     2.4138    14.36   0.000      29.9384    39.40301
    lavggift |   13.10639   .6775479    19.34   0.000     11.77805    14.43474
       _cons |  -55.06357   2.947971   -18.68   0.000    -60.84313   -49.28401
-------------+----------------------------------------------------------------
      /sigma |   24.33349   .4597725                       23.4321    25.23488
------------------------------------------------------------------------------
         2,561  left-censored observations at gift <= 0
         1,707     uncensored observations
             0 right-censored observations
```

```
. margins, dydx(*) predict(ystar(0,.))

Average marginal effects                        Number of obs     =      4,268
Model VCE    : OIM

Expression   : E(gift*|gift>0), predict(ystar(0,.))
dy/dx w.r.t. : 1.resplast weekslast mailsyear propresp lavggift

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  1.resplast |   .2346497   .4746847     0.49   0.621    -.6957153    1.165015
   weekslast |  -.0493058   .0062147    -7.93   0.000    -.0614864   -.0371251
   mailsyear |   .5958669   .2634876     2.26   0.024     .0794406    1.112293
    propresp |    13.3826   .9253948    14.46   0.000     11.56886    15.19634
    lavggift |   5.058959   .2714734    18.64   0.000     4.526881    5.591037
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

```
. margins, dydx(mailsyear) at(mailsyear = (1 2 3)) predict(ystar(0,.)) vsquish

Average marginal effects                        Number of obs     =      4,268
Model VCE    : OIM

Expression   : E(gift*|gift>0), predict(ystar(0,.))
dy/dx w.r.t. : mailsyear
1._at        : mailsyear       =           1
2._at        : mailsyear       =           2
3._at        : mailsyear       =           3

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
mailsyear    |
         _at |
          1  |    .562269   .2338624     2.40   0.016     .1039072    1.020631
          2  |   .5932174   .2611319     2.27   0.023     .0814082    1.105027
          3  |   .6246454   .2890459     2.16   0.031     .0581258    1.191165
------------------------------------------------------------------------------
```

```
. probit respond i.resplast weekslast mailsyear propresp lavggift

Probit regression                               Number of obs     =      4,268
                                                LR chi2(5)        =     989.39
                                                Prob > chi2       =     0.0000
Log likelihood = -2377.6399                     Pseudo R2         =     0.1722

------------------------------------------------------------------------------
     respond |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  1.resplast |   .1255117   .0572287     2.19   0.028     .0133455    .2376779
   weekslast |   -.004517   .0007112    -6.35   0.000    -.0059109   -.0031231
   mailsyear |   .1408493   .0322642     4.37   0.000     .0776125     .204086
    propresp |   1.853517   .1143114    16.21   0.000     1.629471    2.077563
    lavggift |   .0475477   .0318031     1.50   0.135    -.0147851    .1098806
       _cons |  -1.391889    .134267   -10.37   0.000    -1.655047    -1.12873
------------------------------------------------------------------------------

. qui probit respond i.resplast weekslast mailsyear propresp lavggift

. gen g_mailsyear = _b[mailsyear]

. predict xg, xb
```

3. Motivation for Two-Part Models

- Break outcome into two parts:

1. The *participation decision*: $y = 0$ versus $y > 0$.
2. The *amount decision* (how much?)

- In a Tobit model, the partial effects of x_j on

$$P(y > 0|\mathbf{x}) \text{ and } E(y|\mathbf{x}, y > 0)$$

must have the same sign.

- For continuous variables x_j and x_h,

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

- If x_j has twice the effect of 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, there are fixed costs of entering the labor force.
- Perhaps there are fixed psychological costs of buying life insurance.

4. A General Formulation

- s is a binary variable that determines whether y is zero or strictly positive.
- $w^* > 0$ is a 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.
- We observe s because $s = 1$ if and only if $y > 0$:

$$s = 1[y > 0].$$

- w^* is only observed when $s = 1$, in which case $w^* = y$.
- Seems we would want to allow s and w^* to be dependent, but it is not so easy.

- A useful assumption is that s and w^* are independent conditional on explanatory variables $\mathbf{x}$:

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

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

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

- It is enough to have conditional mean independence:

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

- When $s = 1$, we can write the "conditional" expectation as

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

- 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 decision, s, and the latent amount, w^*, even after conditioning on $\mathbf{x}$.
- Unfortunately, these models have been dubbed **selection models**.
- The name has led to confusion about what data is observed for corner solution responses.

- 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 $y > 0$.
- We are trying to get flexible models for $D(y|\mathbf{x})$ that can be given behavioral content.
- There is no missing data problem!

5. Truncated Normal Hurdle Model

- Cragg (1971) proposed a two-part extension of the Type I Tobit model.
- With $y = s \cdot w^*$, $D(s, w^*|\mathbf{x}) = D(s|\mathbf{x})D(w^*|\mathbf{x})$.
- The binary variable s follows a probit model:

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

which is the same as

$$s = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0]$$
$$v|\mathbf{x} \sim Normal(0, 1)$$

- w^* has a **truncated normal distribution** with parameters that vary freely from those in the probit.
- We can write

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

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

xx graph?

- 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 $y > 0$.

- We combine this with

$$P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma}).$$

- Called the **truncated normal hurdle (TNH) model**.
- Nice feature of the TNH model: It reduces to the Type I

Tobit model when $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ (K restrictions).

- 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 $s_i = 1[y_i > 0]$ on $\mathbf{x}_i$.
- The MLEs of $\boldsymbol{\beta}$ and σ (or $\boldsymbol{\beta}$ and σ^2) use the log likelihood

$$\sum_{i=1}^{N} 1[y_i > 0]\{-\log[\Phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma)] + \log(\phi[(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma]) - \log(\sigma)\}.$$

- The latter estimation problem is sometimes called **truncated normal regression**.

• The conditional expectation has the same form as the Type I Tobit:

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

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

• In the TNH model, γ_j/γ_h (ratio of PEs on participation probabilities) can be completely different from β_j/β_h (ratio of PEs on $E(y|\mathbf{x}, y > 0)$).

- The unconditional expectation for the Cragg TNH model is

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

- PE for a continuous x_j:

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

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

- The PE sign is unambiguous if γ_j, β_j have the same sign.

- The estimated average partial (marginal) effect is

$$\widehat{APE}_j = N^{-1}\sum_{i=1}^{N}\left\{\hat{\gamma}_j\phi(\mathbf{x}_i\hat{\boldsymbol{\gamma}})[\mathbf{x}_i\hat{\boldsymbol{\beta}} + \sigma\lambda(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})] + \hat{\beta}_j\Phi(\mathbf{x}_i\hat{\boldsymbol{\gamma}})\theta(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})\right\}$$

- This $\widehat{APE}_j$ can be compared with linear, Tobit.
- Might want to set a key element of **x** to different values, average out the rest.

- 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 $E(y|\mathbf{x})$ with respect to z_j.
- Bootstrapping is convienent for obtaining valid standard errors for APEs.

- Can get goodness-of-fit measures, as with Tobit.
- For example, the squared correlation between y_i and

$$\hat{E}(y_i|\mathbf{x}_i) = \Phi(\mathbf{x}_i\hat{\boldsymbol{\gamma}})\left[\mathbf{x}_i\hat{\boldsymbol{\beta}} + \hat{\sigma}\lambda(\mathbf{x}_i\hat{\boldsymbol{\beta}}/\hat{\sigma})\right].$$

- Or, use a sum-of-squared residuals form:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}[y_i - \hat{E}(y_i|\mathbf{x}_i)]^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

6. Lognormal Hurdle Model

- Cragg (1971) also suggested the lognormal distribution conditional on a positive outcome.
- We can express y as

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

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

- As in the TNH model, 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 PE for continuous x_j:

$$\gamma_j \phi(\mathbf{x}\boldsymbol{\gamma}) \exp(\mathbf{x}\boldsymbol{\beta} + \sigma^2/2) + \beta_j \Phi(\mathbf{x}\boldsymbol{\gamma}) \exp(\mathbf{x}\boldsymbol{\beta} + \sigma^2/2).$$

- Can easily estimate the average of these:

$$N^{-1} \sum_{i=1}^{N} \left[\hat{\gamma}_j \phi(\mathbf{x}_i \hat{\boldsymbol{\gamma}}) \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}} + \hat{\sigma}^2/2) + \hat{\beta}_j \Phi(\mathbf{x}_i \hat{\boldsymbol{\gamma}}) \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}} + \hat{\sigma}^2/2) \right].$$

- This $\widehat{APE}_j$ can be compared with linear, Tobit, TNH, and so on.

- The semi-elasticity of $E(y|\mathbf{x})$ with respect to x_j is obtained by differentiating

$$\log[\Phi(\mathbf{x}\boldsymbol{\gamma})] + \mathbf{x}\boldsymbol{\beta} + \sigma^2/2,$$

which gives

$$\gamma_j\lambda(\mathbf{x}\boldsymbol{\gamma}) + \beta_j,$$

where $\lambda(\cdot)$ is the inverse Mills ratio.

- Multiply by 100 to turn into a percent.
- If $x_j = \log(z_j)$, $\gamma_j\lambda(\mathbf{x}\boldsymbol{\gamma}) + \beta_j$ is the elasticity of $E(y|\mathbf{x})$ with respect to z_j.

- Estimation of the parameters is straightforward.
- The log-likelihood function for a random draw i:

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

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

- As with the truncated normal hurdle model, estimation of the MLE estimates can be obtained in two steps:

1. Probit of s_i on $\mathbf{x}_i$ to estimate $\boldsymbol{\gamma}$.
2. OLS regression of $\log(y_i)$ on $\mathbf{x}_i$ for observations with $y_i > 0$ to estimate $\boldsymbol{\beta}$ and σ^2.

- The MLE of σ^2 can be used or the usual degrees-of-freedom adjusted estimator can be used; both are consistent.

- In computing the log likelihood to compare fit across models, must include the terms $\log(y_i)$.
- The second-part models can be formally compared using Vuong's (1988, *Econometrica*) **model selection statistic**.
- The null is that both models are misspecified but fit equally well.
- The models are nonnested.
- Obtain the usual t statistic from the regression

$$\ell_{i1}(\hat{\boldsymbol{\theta}}_1) - \ell_{i2}(\hat{\boldsymbol{\theta}}_2) \text{ on } 1,\ i = 1, \ldots, N$$

7. Applications of Hurdle Models

```
. use charity

* Lognormal Hurdle:

. reg lgift resplast weekslast mailsyear propresp lavggift

      Source |       SS           df       MS      Number of obs   =     1,707
-------------+----------------------------------   F(5, 1701)      =   1317.26
       Model |  664.151099         5   132.83022   Prob > F        =    0.0000
    Residual |  171.526333     1,701  .100838526   R-squared       =    0.7947
-------------+----------------------------------   Adj R-squared   =    0.7941
       Total |  835.677432     1,706  .489846091   Root MSE        =    .31755

------------------------------------------------------------------------------
       lgift |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    resplast |  -.0090969   .0209414    -0.43   0.664    -.0501706    .0319768
   weekslast |   .0002882   .0003359     0.86   0.391    -.0003706     .000947
   mailsyear |   .0076318   .0122335     0.62   0.533    -.0163625    .0316261
    propresp |   .0474344   .0404902     1.17   0.242    -.0319814    .1268501
    lavggift |   .9161706   .0115061    79.62   0.000      .893603    .9387383
       _cons |   .2042159   .0516533     3.95   0.000     .1029053    .3055266
------------------------------------------------------------------------------
```

```
. gen b_mailsyear = _b[mailsyear]

. predict xb1
(option xb assumed; fitted values)

. predict u1, resid
(2,561 missing values generated)

. di sqrt(e(df_r)/e(N))*e(rmse)
.31699225

. gen sigma_l = sqrt(e(df_r)/e(N))*e(rmse)

. gen sigmasq_l = sigma_l^2

. gen llf_l = log(normalden(u1/sigma_l)) - log(sigma_l) - lgift
(2,561 missing values generated)

. sum llf_l

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       llf_l |      1,707   -2.922409    4.202741  -167.9574  -.6035189

. * Compute log likelihood for the positive part
. gen ln_llf = r(N)*r(mean)

. di ln_llf
-4988.5527
```

```
. * Obtain APE with respect to mailsyear:
. gen pe_mailsyear = g_mailsyear*normalden(xg)*exp(xb1 + sigmasq_l/2) ///
>         + b_mailsyear*normal(xg)*exp(xb1 + sigmasq_l/2)

. sum pe_mailsyear

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
pe_mailsyear |      4,268    .8847765    2.224905   .0360896   136.7779
```

```
. * Truncated Normal Hurdle
.
. truncreg gift resplast weekslast mailsyear propresp lavggift, ll(0)
(note: 2,561 obs. truncated)

Truncated regression
Limit:   lower =          0                     Number of obs     =      1,707
         upper =       +inf                     Wald chi2(5)      =     964.45
Log likelihood = -6069.9612                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
        gift |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    resplast |  -2.605616   1.681491    -1.55   0.121    -5.901277    .6900461
   weekslast |  -.0506036   .0280123    -1.81   0.071    -.1055068    .0042995
   mailsyear |  -.9740504   .9199597    -1.06   0.290    -2.777138    .8290375
    propresp |   3.013063   3.239026     0.93   0.352    -3.335312    9.361437
    lavggift |   39.51495   1.274278    31.01   0.000     37.01741    42.01249
       _cons |  -97.34161   5.583595   -17.43   0.000    -108.2853   -86.39796
-------------+----------------------------------------------------------------
      /sigma |   16.95007   .5022369    33.75   0.000      15.9657    17.93443
------------------------------------------------------------------------------
```

```
. gen sigma_t = _b[/sigma]

. predict xb2, xb

. gen u2 = gift - xb2

. gen llf_t = log(normalden(u2/sigma_t)) ///
>                - log(sigma_t) - log(normal(xb2/sigma_t))

. replace llf_t = . if respond == 0
(2,561 real changes made, 2,561 to missing)

. sum llf_t

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       llf_t |      1,707   -3.555923    2.711403  -88.88192  -1.810982

. * Compute log likelihood for the positive part
. gen tn_llf = r(N)*r(mean)

. di tn_llf
-6069.9614
```

```
. * Compute APE for mailsyear:

. * Compute APE with respect to mailsyear
. gen xb2_sigma_t = xb2/sigma_t

. gen imr_tn = normalden(xb2_sigma_t)/normal(xb2_sigma_t)

. gen theta = 1 - imr_tn*(xb2_sigma_t + imr_tn)

.
. drop pe_mailsyear

. gen pe_mailsyear = g_mailsyear*normalden(xg)*(xb2 + sigma_t*imr_tn) ///
>         + b_mailsyear*normal(xg)*theta

. sum pe_mailsyear

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
pe_mailsyear |      4,268    .6264226    .5893324   .0598148    8.80879

. * APE for Tobit was .596, for LNH .885.
```

```
. * Compute Vuong Statistic:

. gen diff = llf_l - llf_t
(2,561 missing values generated)

. reg diff

      Source |       SS           df       MS      Number of obs   =     1,707
-------------+----------------------------------   F(0, 1706)      =      0.00
       Model |           0         0           .   Prob > F        =         .
    Residual |  11109.9727     1,706   6.5122935   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =    0.0000
       Total |  11109.9727     1,706   6.5122935   Root MSE        =    2.5519

------------------------------------------------------------------------------
        diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .633514   .0617661    10.26   0.000     .5123687    .7546593
------------------------------------------------------------------------------

* Heavily favors lognormal model
```

```
. * Labor Supply Example

. use mroz

. tab inlf

   =1 if in |
  lab frce, |
       1975 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        325       43.16       43.16
          1 |        428       56.84      100.00
------------+-----------------------------------
      Total |        753      100.00

. * Compute Vuong test for truncated normal versus lognormal.

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

```
. 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)
```

```
. * Obtain the MLE of sigma:

. di sqrt(e(df_r)/e(N))*e(rmse)
.88406695

. gen sigma_l = sqrt(e(df_r)/e(N))*e(rmse)

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

. gen llf_l = log(normalden(u1/sigma_l)) - log(sigma_l) - lhours
(325 missing values generated)

. gen llf_l = log(normalden(u1/sigma_l)) - log(sigma_l) - lhours
(325 missing values generated)

. sum llf_l

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       llf_l |       428   -8.162676    .8157579  -12.81951    -6.2637

. di r(N)*r(mean)
-3493.6254

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

```
. 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]
-------------+----------------------------------------------------------------
    nwifeinc |     .15344   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
------------------------------------------------------------------------------

. di _b[/sigma]
850.766

. gen sigma_t = _b[/sigma]
```

```
. predict xb2, xb

. gen u2 = hours - xb2

. predict xb2, xb

. gen u2 = hours - xb2

. gen llf_t = log(normalden(u2/sigma_t))
               - log(sigma_t) - log(normal(xb2/sigma_t))

. replace llf_t = . if inlf == 0
(325 real changes made, 325 to missing)

. sum llf_t

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

. di r(N)*r(mean)
-3390.6476
```

```
. gen diff = llf_t - llf_l
(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 |  203.970312   427  .477682229           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |  203.970312   427  .477682229           Root MSE      =  .69115

------------------------------------------------------------------------------
        diff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2406023   .0334078     7.20   0.000     .1749381    .3062665
------------------------------------------------------------------------------

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

- We could drop the full distributional assumption and assume

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

- Then

$$E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\gamma})\exp(\mathbf{x}\boldsymbol{\beta}).$$

- $\boldsymbol{\gamma}$ is still estimated by probit.
- $\boldsymbol{\beta}$ can be estimated by Poisson or Exponential QMLE using $y_i > 0$.
- Computation of semi-elasticities and elasticities follows from

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

- For example,

$$\frac{\partial \log[\hat{E}(y|\mathbf{x})]}{\partial x_j} = \hat{\gamma}_j \lambda(\mathbf{x}\hat{\boldsymbol{\gamma}}) + \hat{\beta}_j.$$

8. Exponential Type II Tobit Model

- 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 (T2T) model has been applied to missing data problems.
- Here, we use an exponential version of the T2T model as a corner solution model.

- We can still write y as

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

- Because $Var(v) = 1$, $Var(u) = \sigma^2$,

$$Cov(u,v) = \rho\sigma = Corr(u,v) \cdot \sigma.$$

- Compared with the lognormal hurdle model, we have added a correlation parameter, ρ.

- Obtaining the pdf is a bit tricky.

- After tedious calcluations, the log likelihood for a random draw i:

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

- Practical Tip: To use sample selection software for corner solution outcomes, define $\log(y_i)$ as the response variable.
- The data are "missing" on $\log(y_i)$ when $y_i = 0$.

- For a true missing data problem, the last term in the log likelihood, $\log(y_i)$, is not included.
- Inclusion of $\log(y_i)$ clearly 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.

Computing Partial Effects

- 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 $\eta = \rho\sigma$ is the same as $\rho = Corr(u, v)$.

- The partial effects on the unconditional expectation of y are

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

$$+ \beta_j \Phi(\mathbf{x}\boldsymbol{\gamma} + \eta) \exp(\mathbf{x}\boldsymbol{\beta} + \sigma^2/2).$$

- Very similar to LNH, which sets $\eta = 0 = \rho$.
- The semi-elasticity is

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

Identification without an Exclusion Restriction

• Technically, the ET2T model contains the lognormal hurdle model as a special case ($\rho = 0$).

• 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

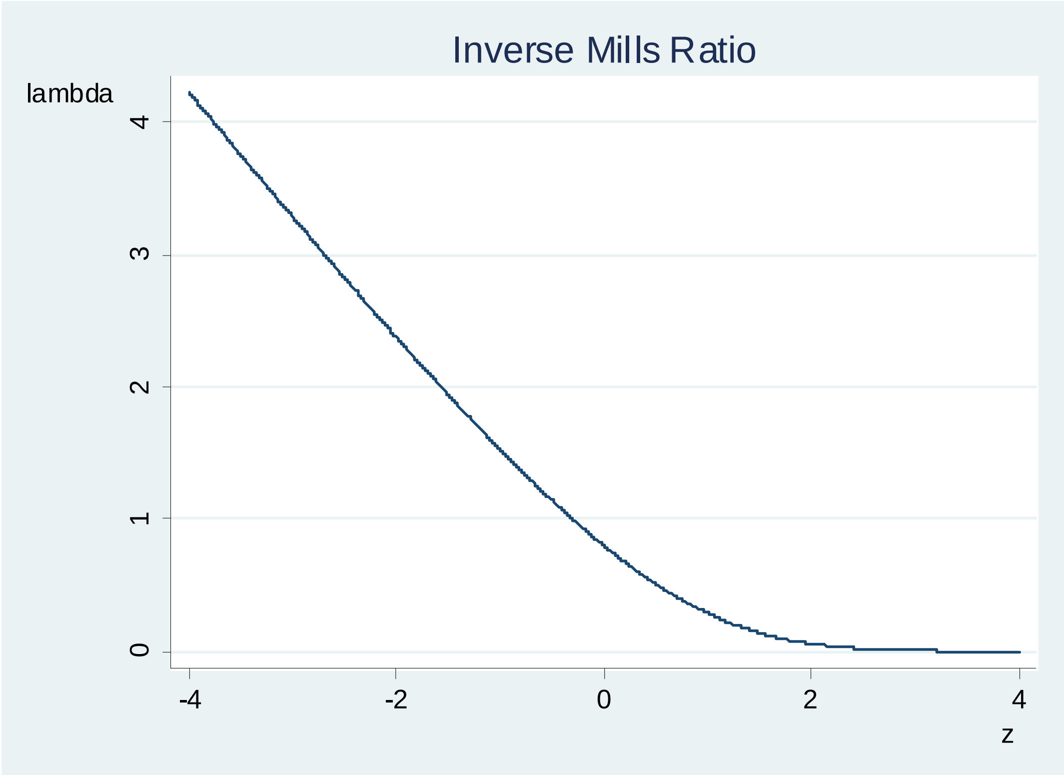
$$s = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0].$$

- Various ways to see the potential problem.
- First, it can be shown that

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

where $\eta = \rho\sigma$.

- $\boldsymbol{\gamma}$ is identified by probit, so we can treat it as known.
- $E[\log(y)|\mathbf{x}, y > 0]$ nominally identifies $\boldsymbol{\beta}$ and η only because $\lambda(\cdot)$ is a nonlinear function.
- But 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

- In the ET2T 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 except for the presence of $\eta = \rho\sigma$.

- Because **x** always should include a constant, η is not separately identified by $E(y|\mathbf{x})$ (niether 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.

- Technically, the parameters are identified without an exclusion restriction, so we can try to estimate the full model with the same vector **x** appearing in the participation and amount equations.
- In practice, it rarely works well. For example, it can give unrealistic estimates of ρ (such as close to 1 or –1).
- One should view estimating the ET2T without an exclusion restriction with extreme skepticism – as in any case where identification is due to a nonlinearity.

```
. use charity

. heckman lgift resplast weekslast mailsyear propresp lavggift, ///
>         select(respond = resplast weekslast mailsyear propresp lavggift)

Heckman selection model                         Number of obs     =      4,268
(regression model with sample selection)        Censored obs      =      2,561
                                                Uncensored obs    =      1,707

                                                Wald chi2(5)      =    5474.35
Log likelihood =   -2802.1                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lgift        |
    resplast |  -.0219504   .0229974    -0.95   0.340    -.0670245    .0231238
   weekslast |   .0016271   .0003569     4.56   0.000     .0009276    .0023266
   mailsyear |    -.01567   .0133885    -1.17   0.242     -.041911    .0105709
    propresp |  -.2927994   .0481844    -6.08   0.000     -.387239   -.1983597
    lavggift |    .913751    .012709    71.90   0.000     .8888419    .9386601
       _cons |   .6619247   .0606737    10.91   0.000     .5430064     .780843
-------------+----------------------------------------------------------------
```

```
respond      |
    resplast |   .1265888   .0564515     2.24   0.025     .0159459    .2372316
   weekslast |  -.0041173   .0007034    -5.85   0.000    -.005496   -.0027386
   mailsyear |   .1243341   .0318364     3.91   0.000     .0619358    .1867323
    propresp |    1.80489   .1129476    15.98   0.000     1.583517    2.026263
    lavggift |   .0851231   .0315661     2.70   0.007     .0232547    .1469915
       _cons |  -1.468231   .1329752   -11.04   0.000    -1.728858   -1.207604
-------------+----------------------------------------------------------------
     /athrho |  -1.023908   .0577399   -17.73   0.000    -1.137076   -.9107401
    /lnsigma |  -.9382428   .0265996   -35.27   0.000    -.9903771   -.8861085
-------------+----------------------------------------------------------------
         rho |  -.7714536   .0233765                     -.8134273   -.7214873
       sigma |   .3913148   .0104088                      .3714366    .4122569
      lambda |  -.3018813   .0161214                     -.3334785    -.270284
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    73.07   Prob > chi2 = 0.0000

. sum lgift

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       lgift |      1,707    2.652349    .6998901   .6931472   5.521461

. * Compute log likelihood for "selection" model, both parts
. di e(ll) - r(N)*r(mean)
-7329.6597
```

```
. * Do Heckman two-step:

. heckman lgift resplast weekslast mailsyear propresp lavggift, ///
>    select(respond = resplast weekslast mailsyear propresp lavggift) twostep

Heckman selection model -- two-step estimates   Number of obs     =      4,268
(regression model with sample selection)        Censored obs      =      2,561
                                                Uncensored obs    =      1,707

                                                Wald chi2(5)      =    4730.55
                                                Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lgift        |
    resplast |  -.0305503   .0273419    -1.12   0.264    -.0841394    .0230388
   weekslast |    .001681   .0008859     1.90   0.058    -.0000554    .0034175
   mailsyear |  -.0205983   .0218647    -0.94   0.346    -.0634523    .0222558
    propresp |  -.3738332   .2487012    -1.50   0.133    -.8612787    .1136123
    lavggift |   .9039068   .0153245    58.98   0.000     .8738714    .9339423
       _cons |   .8126557    .356659     2.28   0.023     .1136169    1.511694
-------------+----------------------------------------------------------------
```

respond						
resplast	.1255118	.0572286	2.19	0.028	.0133457	.2376779
weekslast	-.004517	.0007112	-6.35	0.000	-.0059109	-.0031231
mailsyear	.1408493	.0322642	4.37	0.000	.0776126	.204086
propresp	1.853517	.1143113	16.21	0.000	1.629471	2.077563
lavggift	.0475477	.031803	1.50	0.135	-.014785	.1098805
_cons	-1.391889	.1342668	-10.37	0.000	-1.655047	-1.128731
mills						
lambda	-.3770068	.2165736	-1.74	0.082	-.8014833	.0474697
rho	-0.87463					
sigma	.43104751					

```
. * "By hand":

. reg lgift resplast weekslast mailsyear propresp lavggift imr if respond

      Source |       SS           df       MS      Number of obs   =     1,707
-------------+----------------------------------   F(6, 1700)      =   1099.62
       Model |  664.467637         6  110.744606   Prob > F        =    0.0000
    Residual |  171.209796     1,700  .100711644   R-squared       =    0.7951
-------------+----------------------------------   Adj R-squared   =    0.7944
       Total |  835.677432     1,706  .489846091   Root MSE        =    .31735

------------------------------------------------------------------------------
       lgift |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    resplast |  -.0305503   .0241749    -1.26   0.207     -.077966    .0168655
   weekslast |    .001681   .0008544     1.97   0.049     5.34e-06    .0033567
   mailsyear |  -.0205983   .0200756    -1.03   0.305    -.0599737    .0187772
    propresp |  -.3738335   .2410423    -1.55   0.121    -.8466042    .0989373
    lavggift |   .9039068   .0134193    67.36   0.000     .8775868    .9302269
         imr |   -.377007   .2126556    -1.77   0.076    -.7941013    .0400873
       _cons |    .812656   .3470589     2.34   0.019     .1319485    1.493363
------------------------------------------------------------------------------
```

```
. qui probit respond resplast weekslast mailsyear propresp lavggift

. gen probit_llf = e(ll)

. gen tnh_llf = probit_llf + tn_llf

. di tnh_llf
-8447.6016

* LLF for Tobit is -9115.3322.

.
. gen lnh_llf = probit_llf + ln_llf

. di lnh_llf
-7366.1924
```

```
. use mroz

. 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
------------------------------------------------------------------------------

. * rho = -.972 is not plausible.
```

```
. di e(ll)
-938.8208

. sum lhours

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

. di e(ll) - r(N)*r(mean)
-3877.8798

. * This value of the LLF is well below that of the truncated normal
. * hurdle model, which is -3,791.95 = -401.302 - 3390.648.
. * (And the TNH has one fewer parameter.)
. * Of course, -3,877.88 is above that for the lognormal hurdle model
. * (-3,894.93) because the ET2T model nests the LNH model.

. * But with such a silly estimate of rho, the ET2T estimates should
. * not be trusted.
```

Estimation with an Exclusion Restriction

• The ET2T is much more convincing when the covariates determining the amount are a strict subset of those affecting participation.

• 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 $\mathbf{x}$ and $\mathbf{x}_1$ contain unity but $\mathbf{x}_1$ is a strict subset of $\mathbf{x}$.

• If we write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ and $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')$ then we are assuming $\boldsymbol{\gamma}_2 \neq \mathbf{0}$, so that at least one element of $\mathbf{x}_2$ affects participation.

• 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 much are better identified because $\lambda(\mathbf{x}\boldsymbol{\gamma})$ is not an exact function of $\mathbf{x}_1$.

• There is never a need to excluse variables from the participation equation. Doing so can cause inconsistency.

- Exclusion restrictions can be hard to come by.
- Need something affecting the fixed cost of participating but not affecting the amount of y.
- What would affect a firm's decision to grant of equity incentives that would arguably not affect the amount?
- What would affect a family's decision to contribute to charity but not the amount?

- Given an exclusion restriction, can use Vuong's test to compare the ET2T model and the lognormal hurdle model (no exclusion restriction but $\rho = 0$).
- With one exclusion restriction, the models have the same number of parameters.

Using y in the Selection Model

- It is a mistake to use y rather than $\log(y)$ in the linear amount equation.
- We would then write

$$y = 1[\mathbf{x}\boldsymbol{\gamma} + v > 0] \cdot (\mathbf{x}_1\boldsymbol{\beta}_1 + u)$$

where (u, v) is bivariate normal *independent* of $\mathbf{x}$.

- Then

$$\mathbf{x}_1\boldsymbol{\beta}_1 + u < 0$$
$$\mathbf{x}\boldsymbol{\gamma} + v > 0$$

can both occur, which means $y < 0$.

- When $y \geq 0$ with a corner at zero, it only makes sense to apply the Type Two Tobit model to $\log(y)$.

```
. * Apply the two-step Heckman estimator to gift rather than lgift.

. * Gives many negative predictions conditional on gift > 0!

. reg gift resplast weekslast mailsyear propresp lavggift imr if respond

      Source |       SS           df       MS      Number of obs   =     1,707
-------------+----------------------------------   F(6, 1700)      =    338.05
       Model |  333613.342         6  55602.2236   Prob > F        =    0.0000
    Residual |  279617.474     1,700  164.480867   R-squared       =    0.5440
-------------+----------------------------------   Adj R-squared   =    0.5424
       Total |  613230.815     1,706  359.455343   Root MSE        =    12.825

------------------------------------------------------------------------------
        gift |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    resplast |  -1.194293   .9769745    -1.22   0.222    -3.110492     .721906
   weekslast |  -.0192372   .0345268    -0.56   0.577    -.0869567    .0484823
   mailsyear |  -1.364588   .8113086    -1.68   0.093    -2.955856    .2266811
    propresp |  -.6846713   9.741169    -0.07   0.944    -19.79061    18.42127
    lavggift |    20.7177   .5423092    38.20   0.000     19.65404    21.78136
         imr |  -1.150011   8.593988    -0.13   0.894    -18.00592     15.7059
       _cons |  -29.91137   14.02558    -2.13   0.033     -57.4206   -2.402144
------------------------------------------------------------------------------

. predict gift_h
(option xb assumed; fitted values)

. count if gift_h < 0
  439
```