The Tobit Model

Econometrics II

Ricardo Mora

Department of Economics
Universidad Carlos III de Madrid
Máster Universitario en Desarrollo y Crecimiento Económico

Notes

Outline

Notes

Basic Setup

Utility Function

- $U = U(C, L)$
- C: consumption
- L: leisure

Marginal Utility of Consumption and Leisure

- $U_C = \left.\frac{\partial U}{\partial C}\right|_L > 0$, $\left.\frac{\partial U_C}{\partial C}\right|_L < 0$: more consumption gives more utility at a decreasing rate
- $U_L = \left.\frac{\partial U}{\partial L}\right|_C > 0$, $\left.\frac{\partial U_L}{\partial L}\right|_C < 0$: additional leisure gives additional utility at a decreasing rate

Notes

Marginal Rate of Substitution in Consumption

Marginal Rate of Substitution: the individual's value of leisure

- $MRS = \left.\frac{\partial C}{\partial L}\right|_U = -\frac{U_L}{U_C}$
- (By how much I can reduce my consumption without losing utility if I increase my leisure)

Cobb-Douglas: $U = C^\alpha L^\beta \rightarrow MRS = \left(\frac{\alpha}{\beta}\right)\left(\frac{C}{L}\right)$

- Increasing in consumption
- Decreasing in leisure

Notes

Time and Budget Constraints

Time constraint: $L + h = T$

- h: hours of work
- T: total hours available

Budget constraint: $C = w * h + V$

- w: hourly wage (the opportunity cost of one unit of leisure)
- V: non-labour income

$C + wL = wT + V$

- wL : total cost of leisure
- $wT + V$: time and non-labor income

Notes

The Optimal Allocation of Leisure

$\max\ U(C,L)$ s.t. $C + wL = wT + V$

- Internal Solution: $MRS = w$
- the value of leisure equals its cost

- $MRS > w$: a small increase in leisure will increase utility
- $MRS < w$: a small increase in work will increase utility (via higher consumption)

Notes

The Reservation Wage

Individuals work if the wage is larger than their reservation wage

- $w_R = MRS(T, V)$
- For any $w > w_R$: Internal Solution ($h > 0$)
- For any $w \leq w_R$: Corner solution ($h = 0$)

- The higher the market wage, the more likely are individuals to work
- The reservation wage depends on non-labour income and on the individual's preferences on leisure and consumption

Notes

A Two Stage Procedure

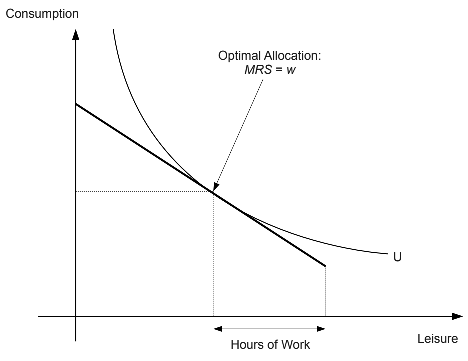
Hours worked decision can be decomposed into two stages

- first decision: participation decision: $w > w_R$
- second decision: if $w > w_R$, how many working hours?

- the first decision is like a Probit model because the participation decision is binary
- the second decision is like a linear regression model because the amount of time worked can be considered continuous
- both decisions are strongly linked: factors that make a married woman more likely to participate, tend to make her work more hours

Notes

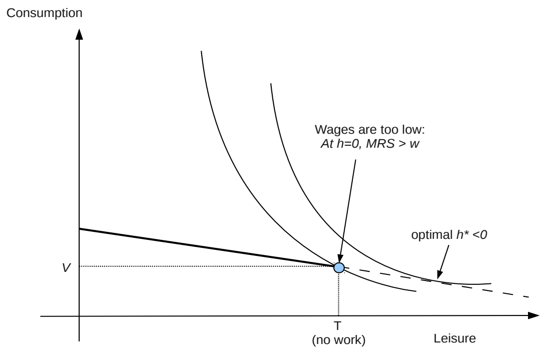
The Optimal Allocation of Leisure: Internal Solution



internal solution: $h = h^* (MRS = w) > 0$

Notes

The Optimal Allocation of Leisure: Corner Solution



corner solution: $h = 0$ if $h^* (MRS = w) \leq 0$

Notes

The Tobit Model

Example: Married Women Labor Supply

- optimality condition $(MRS = w)$: $h^* = \beta x + \varepsilon, \varepsilon \sim N(0, \sigma^2)$
- participation condition $(MRS < w)$: $h^* > 0$
- If $h^* > 0$, then actual hours of work: $h = h^*$
- If $h^* \leq 0$, then actual hours of work: $h = 0$

$$h = \max\{0, \beta x + \varepsilon\}, \varepsilon \sim N(0, \sigma^2)$$

Notes

Labor Suppy Controls

Which controls should be in vector x?

- Personal: Non-labor income, spouse's income, number of kids, human capital,...
- Economic conditions: market wages, unemployment rates,...
- Strictly speaking, for the labor supply we require the wage offers. This creates two problems:
  - We do not have information on wage offers for those who are not working.
  - A worker's wage offer is likely related to unobservable characteristics which arguably affect simultaneously the worker's labor supply: Wages and hours worked are simultaneously determined for each worker.

Notes

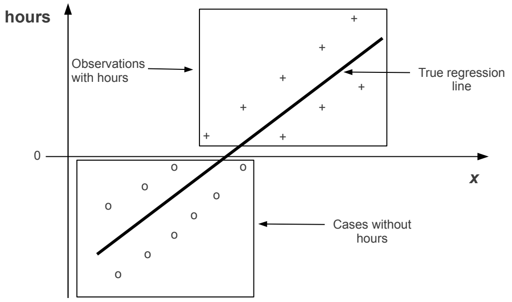
Observable Data

- the econometrician observes whether the married woman participates in the labor market or not
- if the married woman participates, then the econometrician observes the hours of work
- if the married woman does not participate, the econometrician does not observe the optimal number of hours that the married woman would choose to work (in this case, it would be a negative number)

Notes

Using Only the Married Women Who are Working

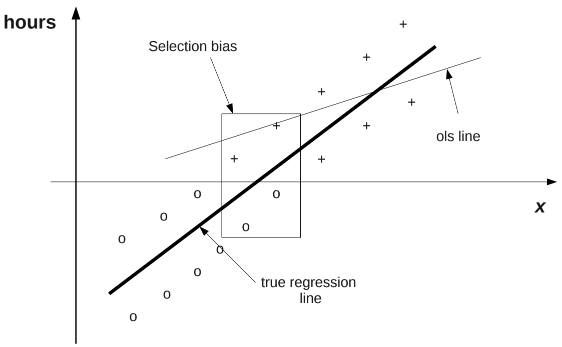
Can we estimate β by OLS using only the data from the married women who choose to work?



Notes

Selection Bias

The OLS sample is not iid: we only observe (h_i, x_i) if $h_i > 0$



Notes

ML Estimation (1/2)

If we estimate by Maximum Likelihood, we use the full sample: including women who choose to work with information of the hours they work and also women who choose not to work

Density of a woman who works $h_i > 0$ hours

$$\begin{aligned} f(h_i|x_i) &= f(\beta_0 x_i + \varepsilon_i|x_i) \\ &= \left(\frac{1}{\sigma_0}\right)\phi\left(\frac{\varepsilon_i}{\sigma_0}\right) \end{aligned}$$

Probability that a woman does not work $(h_i = 0)$

$$\begin{aligned} \Pr(h_i = 0|x_i) &= \Pr(\beta_0 x_i + \varepsilon_i \leq 0|x_i) \\ &= 1 - \Phi\left(\frac{\beta_0 x_i}{\sigma_0}\right) \end{aligned}$$

Notes

ML Estimation (2/2)

- Writing both cases simultaneously:

$$f(h_i|x_i) = \left[\left(\frac{1}{\sigma_0}\right)\phi\left(\frac{h_i-\beta_0 x_i}{\sigma_0}\right)\right]^{1(h_i>0)}\left[1-\Phi\left(\frac{\beta_0 x_i}{\sigma_0}\right)\right]^{1(h_i=0)}$$

Log-likelihood for observation i

$$\begin{aligned} l_i(\beta,\sigma) = \; & 1(h_i>0)\log\left(\left(\tfrac{1}{\sigma}\right)\phi\left(\tfrac{h_i-\beta x_i}{\sigma}\right)\right) \\ & +1(h_i=0)\log\left(1-\Phi\left(\tfrac{\beta x_i}{\sigma}\right)\right) \end{aligned}$$

Notes

ML Estimation

The Tobit Model

- $h^* = \beta x + \varepsilon$
- $\varepsilon \sim N(0,\sigma^2)$
- $\begin{cases} \text{if } h^*>0 \Rightarrow h=\beta x+\varepsilon \\ \text{if } h^*\leq 0 \Rightarrow h=0 \end{cases}$

$$\begin{aligned} \hat{\beta}^{ML} = \arg\max \; & \sum_i \left\{1(h_i>0)\log\left(\left(\tfrac{1}{\sigma}\right)\phi\left(\tfrac{h_i-\beta x_i}{\sigma}\right)\right)\right. \\ & \left. +1(h_i=0)\log\left(1-\Phi\left(\tfrac{\beta x_i}{\sigma}\right)\right)\right\} \end{aligned}$$

Notes

Tobit Estimation in Stata

- `tobit`: computes Maximum Likelihood tobit estimation
- margins (`mfx`): marginal means, predictive margins, marginal effects, and average marginal effects
- `test`: Wald tests of simple and composite linear hypothesis
- `lincom`: point estimates, standard errors, testing, and inference for linear combinations of coefficients
- `predict`: predictions, residuals, influence statistics, and other diagnostic measures
- `e(ll)`: returns the log-likelihood for the last estimated model

Notes

`tobit` *depvar* *indvars* `[if] [in] [weight], ll(#) ul(#)`

- tobit fits a model of *depvar* on *indvars* where the censoring values are fixed.
- ll(#) ul(#): left-censoring and right-censoring limits. You must specify at least one of them.
- the usual post estimation commands are available

Notes

Example: Simulated Data

The Tobit Model

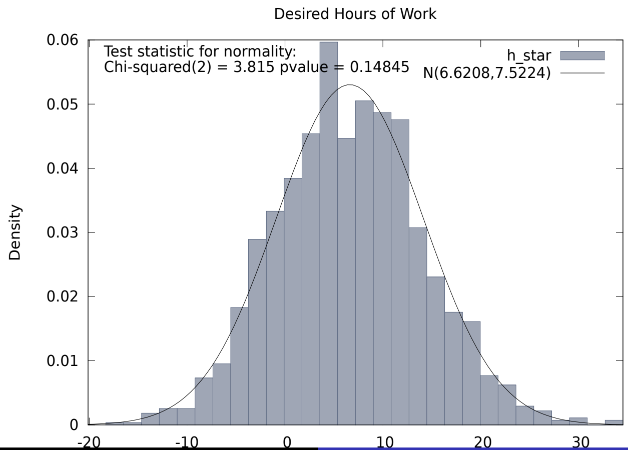
- $h^* = 10 + 0.5 * educ - 5 * kids + \varepsilon$
- $\varepsilon \sim N(0, 49)$

- education makes you willing to work more
- having a kid makes you willing to work less
- $\beta x = 5 + 0.5 * educ - 5 * kids$
- $\sigma = 7$

Notes

Histogram of Desired Hours of Work

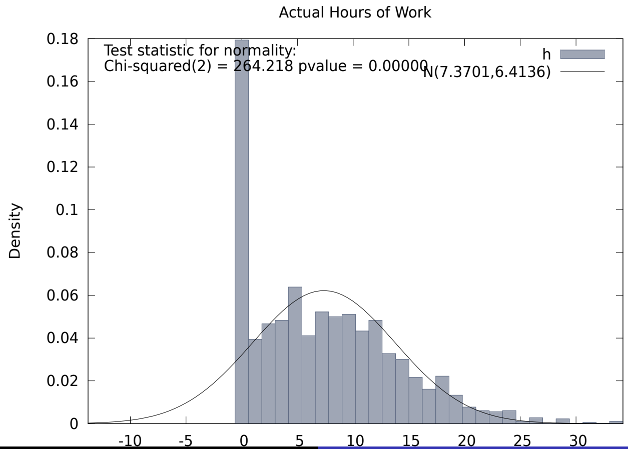
$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$



Notes

Censoring in the Tobit Model

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$



Notes

ols with the Full Sample

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$

```
 regress h educ kids

      Source |       SS       df       MS              Number of obs =    5000
-------------+------------------------------           F(  2,  4997) =  255.70
       Model |  18776.6722     2  9388.33609           Prob > F      =  0.0000
    Residual |   183474.37  4997  36.7169042           R-squared     =  0.0928
-------------+------------------------------           Adj R-squared =  0.0925
       Total |  202251.042  4999  40.4583001           Root MSE      =  6.0594

------------------------------------------------------------------------------
           h |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .4385609   .0417266    10.51   0.000     .3567584    .5203634
        kids |  -4.252116   .2126502   -20.00   0.000    -4.669004   -3.835228
       _cons |   5.932443   .4148881    14.30   0.000      5.11908    6.745806
------------------------------------------------------------------------------

. test (educ=0.5) (kids=-5)

 ( 1)  educ = .5
 ( 2)  kids = -5

       F(  2,  4997) =    9.37
            Prob > F =    0.0001
```


Notes

ols with the Restricted Sample

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$

```
 regress h educ kids if h>0

      Source |       SS       df       MS              Number of obs =    4099
-------------+------------------------------           F(  2,  4096) =  142.99
       Model |  9067.55487     2  4533.77744           Prob > F      =  0.0000
    Residual |  129867.159  4096  31.7058493           R-squared     =  0.0653
-------------+------------------------------           Adj R-squared =  0.0648
       Total |  138934.714  4098  33.9030536           Root MSE      =  5.6308

------------------------------------------------------------------------------
           h |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .3462022   .0426218     8.12   0.000     .2626403    .4297641
        kids |  -3.298671   .2185348   -15.09   0.000    -3.727118   -2.870224
       _cons |    7.74949   .4245528    18.25   0.000     6.917136    8.581844
------------------------------------------------------------------------------

. test (educ=0.5) (kids=-5)

 ( 1)  educ = .5
 ( 2)  kids = -5

       F(  2,  4096) =   42.60
            Prob > F =    0.0000
```


Notes

The Married Women Labor Supply Model
ML Estimation for the Tobit Model
Tobit in Stata
Marginal Effects
Summary

tobit Output

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$

```
 tobit h educ kids, ll(0)

Tobit regression                                  Number of obs   =       5000
                                                  LR chi2(2)      =     491.98
                                                  Prob > chi2     =     0.0000
Log likelihood = -14768.522                       Pseudo R2       =     0.0164

------------------------------------------------------------------------------
           h |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5235172   .0499384    10.48   0.000      .425616    .6214185
        kids |  -5.098043   .2554521   -19.96   0.000    -5.598841   -4.597245
       _cons |   4.905816   .4966346     9.88   0.000     3.932195    5.879438
-------------+----------------------------------------------------------------
      /sigma |   7.111859   .0819006                      6.951298     7.27242
------------------------------------------------------------------------------
  Obs. summary:        901  left-censored observations at h<=0
                      4099     uncensored observations
                         0 right-censored observations

. test (educ=0.5) (kids=-5)

 ( 1)  [model]educ = .5
 ( 2)  [model]kids = -5

       F(  2,  4998) =    0.11
            Prob > F =    0.8950
```


Notes

Predicting Actual Hours of Work for Those who Work

computing $\hat{h}_i^*$ and $\hat{h}_i$

- $\hat{h}_i^*$: predict h_star_hat, xb
- for each observation, $\hat{h}_i = \max\left\{0, \hat{\beta} x_i\right\}$

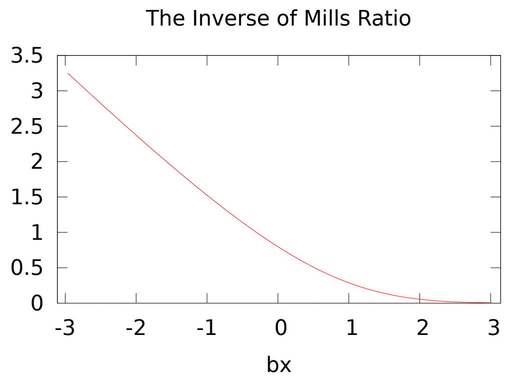
$E\left[h\left|h>0,x\right.\right]$

- $E\left[h\left|h>0,x\right.\right] = \beta x + E\left[\varepsilon\left|\beta x+\varepsilon>0,x\right.\right]$
- it can be shown that: $E\left[h\left|h>0,x\right.\right] = \beta x + \sigma\frac{\phi(\beta x)}{\Phi(\beta x)}$

- $\frac{\phi(\beta x)}{\Phi(\beta x)}$ is the inverse of Mill's ratio

Notes

The Inverse of Mill's Ratio



the higher βx, the higher the probability of participation and the lower the correction

Notes

Predicting Actual Hours of Work

$E[h|x]$

- $E[h|x] = \Pr(h > 0)\, E[h|h > 0, x]$
- it can be shown that: $E[h|x] = \Phi\left(\frac{\beta x}{\sigma}\right)\left[\beta x + \sigma \frac{\phi(\beta x)}{\Phi(\beta x)}\right]$

Notes

Understanding the Coefficients and the Slopes

- the Tobit estimates for the coefficients, $\hat{\beta}$, give the marginal effects on the desired number of hours
- frequently, we also want an estimate of the marginal effects on the probability of working and on the actual hours worked

Notes

Algebraic Marginal Effects

Probability to Participate

- $\frac{\partial \Pr(h_i>0)}{\partial x_j} = \phi(\frac{\beta x}{\sigma})\left(\frac{\beta_j}{\sigma}\right)$

Actual Hours Worked

- $\frac{\partial E(h_i|x)}{\partial x_j} = \beta_j \Phi\left(\frac{\beta x}{\sigma}\right)$
- approx. estimates of this effect can be obtained using OLS over the full sample

Notes

Individual Marginal Effects: Discrete Change

- we may want to get individual marginal effects

Discrete change

- predict index functions $\hat{\beta}^{ML} x_0$ and $\hat{\beta}^{ML} x_1$
- simulate censuring
- generate the individual marginal effects

Notes

Example: The Effect of Having an Extra Kid

```
. // estimate the marginal effect on average actual hours worked of an extra child
. predict x0b, xb
. replace kids=kids+1
(5000 real changes made)
. predict x1b, xb
. gen Mg_kid=(x1b>0)*x1b - (x0b>0)*x0b
. su Mg_kid

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      Mg_kid |      5000    -4.78497    .7293877  -5.098043  -3.086006
```


Notes

Summary

- The Tobit model is like a mixture of the regression model and the Probit model
  - it is partly a Probit model because the participation decision is binary
  - it is partly a linear regression model because among those who work the hours worked can be considered continuous
- Estimating the model by OLS using those who choose to work will usually result in inconsistency because the selected sample is not *iid* (selection bias)
- The Tobit model can be estimated consistently by ML in Stata
- the Tobit model identifies how each control affects both the probability of not censoring and the expectation of the dependent variable given that it is observed

Notes