ADVANCED ECONOMETRICS I

Theory (3/3)

Instructor: Joaquim J. S. Ramalho

E.mail: jjsro@iscte-iul.pt

Personal Website: http://home.iscte-iul.pt/~jjsro

Office: D5.10

Course Website: https://jjsramalho.wixsite.com/advecoi

Fénix: https://fenix.iscte-iul.pt/disciplinas/03089

Ordered choices:

- Values for the dependent variable:

$$Y \in \{0,1,\dots,M-1\}$$

- Latent model:

$$Y_i^* = x_i'\beta + u_i$$

  - x_i cannot include an intercept

- Individual behaviour observed only by intervals:

$$Y_i = \begin{cases} 0 & \text{if } Y_i^* \le \gamma_0 \\ m & \text{if } \gamma_{m-1} < Y_i^* \le \gamma_m, \qquad 1 \le m \le M-2 \\ M-1 & \text{if } Y_i^* > \gamma_{M-2} \end{cases}$$

  - Example:
    - Y_i^* is a latent measure of the health status
    - Y_i is an observed health indicator: poor, satisfactory, good, excellent

- Assumption: the γ_j's are not known

Probabilities:

- Aim:
  - Modelling the probability of observing Y_i^* in a given interval
- Each probability is based on the same $G(\cdot)$ functions used with binary choices, being given by:

$$\begin{aligned} Pr(Y_i = m|x_i) &= Pr(\gamma_{m-1} < Y_i^* \leq \gamma_m|x_i) \\ &= Pr(Y_i^* \leq \gamma_m|x_i) - Pr(Y_i^* < \gamma_{m-1}|x_i) \\ &= Pr(x_i'\beta + u_i \leq \gamma_m|x_i) - Pr(x_i'\beta + u_i < \gamma_{m-1}|x_i) \\ &= Pr(u_i \leq \gamma_m - x_i'\beta|x_i) - Pr(u_i < \gamma_{m-1} - x_i'\beta|x_i) \\ &= G(\gamma_m - x_i'\beta) - G(\gamma_{m-1} - x_i'\beta) \end{aligned}$$

- Hence, the general case is:

$$Pr(Y_i = m|x_i) = \begin{cases} G(\gamma_0 - x_i'\beta) & \text{if } m = 0 \\ G(\gamma_m - x_i'\beta) - G(\gamma_{m-1} - x_i'\beta) & \text{if } 1 \leq m \leq M-2 \\ 1 - G(\gamma_{M-2} - x_i'\beta) & \text{if } m = M-1 \end{cases}$$

Estimation:

- Parameters to be estimated:
  - β
  - $\gamma_0, \dots, \gamma_{M-2}$

- Estimation method:
  - Maximum likelihood

- Most common models:
  - Ordered logit
  - Ordered probit

Stata
ologit $Y X_1 \dots X_k$
oprobit $Y X_1 \dots X_k$

Partial effects:

- Each X_j affects M probabilities:

$$\Delta X_j = 1 \Longrightarrow \Delta Pr(Y = m|X)$$
$$= \begin{cases} -\beta_j g(\gamma_0 - x_i'\beta) & \text{if } m = 0 \\ \beta_j [g(\gamma_m - x_i'\beta) - g(\gamma_{m-1} - x_i'\beta)] & \text{if } 1 \leq m \leq M-2 \\ \beta_j g(\gamma_{M-2} - x_i'\beta) & \text{if } m = M-1 \end{cases}$$

- The sign of β_j is informative about the direction of $\Delta Pr(Y = 0|X)$ and $\Delta Pr(Y = M - 1|X)$ but not of the changes in the remaining probabilities

Multinomial choices:

- Values for the dependent variable:

$$Y \in \{0,1,\ldots,M-1\}$$

- Latent model:
  - Each individual has a given utility associated with each alternative:

$$U_{im} = x'_{im}\beta + u_{im}$$

  - The selected alternative is the one that maximizes utility:

$$Pr(Y_i = m|x_i) = Pr[U_{im} = max(U_{i1},\ldots U_{iM})|x_i]$$

- Main models:
  - Multinomial Logit: $U_{im} \sim Gumbel$ and U_{im} independent $\forall m$
  - Multinomial Probit: $U_{im} \sim Normal$
  - Nested Logit
  - Random Parameters Logit

Explanatory variables:

- x_{im} may include:
  - x_{im}: variables that are different across individuals and alternatives
  - x_m: variables that differ across alternatives but not individuals
  - x_i: variables that differ across individuals but not alternatives

- Example:
  - Y_i - selected means of transport to go to work
  - x_{im} - time that each individual i takes in going to work when using transport m
  - x_m - price of transport m
  - x_i - age of individual i

Multinomial logit:

$$Pr(Y_i = m|x_{im}) = G_m(x'_{im}\beta + x'_i\beta_m) = \frac{e^{x'_{im}\beta + x'_i\beta_m}}{\sum_{j=0}^{M-1} e^{x'_{ij}\beta + x'_i\beta_j}}$$

- β_m has to be normalized, that is for one alternative (base outcome) its value is set to zero
- β cannot include a constant term
- Independence of Irrelevant Alternatives (IIA) – the odds ratio between two alternatives does not depend on the remaining alternatives:

$$\frac{Pr(Y_i = m|x_{im})}{Pr(Y_i = l|x_{im})} = \frac{e^{x'_{im}\beta + x'_i\beta_m}}{e^{x'_{il}\beta + x'_i\beta_l}}$$

- When all explanatory variables are of the type x_{im} and x_m, the choice between alternatives m and l is fully explained by diferences in the alternative characteristics:

$$\frac{Pr(Y_i = m|x_{im})}{Pr(Y_i = l|x_{il})} = \frac{e^{x'_{im}\beta}}{e^{x'_{il}\beta}} = e^{(x'_{im} - x'_{il})\beta}$$

  - Is this case, the model is often called 'conditional logit'

Stata
asclogit Y X_{1m} ..., case(*id*) alternatives(*varname*) casevars(X_i ...) basealternative(*name*)

- When all explanatory variables are of the type x_i, the choice between alternatives m e l is fully explained by diferences between β_m e β_l:

$$\frac{Pr(Y_i = m|x_i)}{Pr(Y_i = l|x_i)} = \frac{e^{x'_i\beta_m}}{e^{x'_i\beta_l}} = e^{x'_i(\beta_m - \beta_l)}$$

Stata
mlogit $Y X_1$ ... X_k, baseoutcome(*0*)

- Estimation:
  - Maximum likelihood based on the following log-likelihood function:
$$LL = \sum_{i=1}^{N} d_{im} log[G_m(x'_{im}\beta + x'_i\beta_m)]$$
  - $d_{im} = 1$ if individual i chooses alternative m

- Partial effects:
  - $\Delta X_{ij} = 1 \Longrightarrow$
    - $\Delta Pr(Y_i = m|X) = \beta_j G_m(\cdot)[d_{im} - G_m(\cdot)]$
    - β_j gives the sign of the partial effect
  - $\Delta X_i = 1 \Longrightarrow$
    - $\Delta Pr(Y_i = m|X) = G_m(\cdot)\left(\beta_j - \bar{\beta}\right)$, onde $\bar{\beta} = \sum_{m=1}^{M-1} \beta_m G_m(\cdot)$
    - β_j gives the sign of the partial effect relative to the base alternative, not the sign of the overall effect

- Testing IIA
  - Hausman test comparing:
    - Full multinomial logit model
    - Multinomial logit model excluding one or more alternatives
  - If multinomial logit is the correct model, then both models produce consistent estimators (null hypothesis)
  - If multinomial logit is not the correct model, then the results generated by both models will be different (alternative hypothesis)

Stata

mlogit $Y X_1$ ... X_k, baseoutcome(*0*) (ou asclogit...)
estimates store *Mod1*
mlogit $Y X_1$ ... X_k if Y!= *3*, baseoutcome(*0*) (ou asclogit...)
estimates store *Mod2*
hausman *Mod1 Mod2*

Multinomial probit:

- Not affected by the IIA property

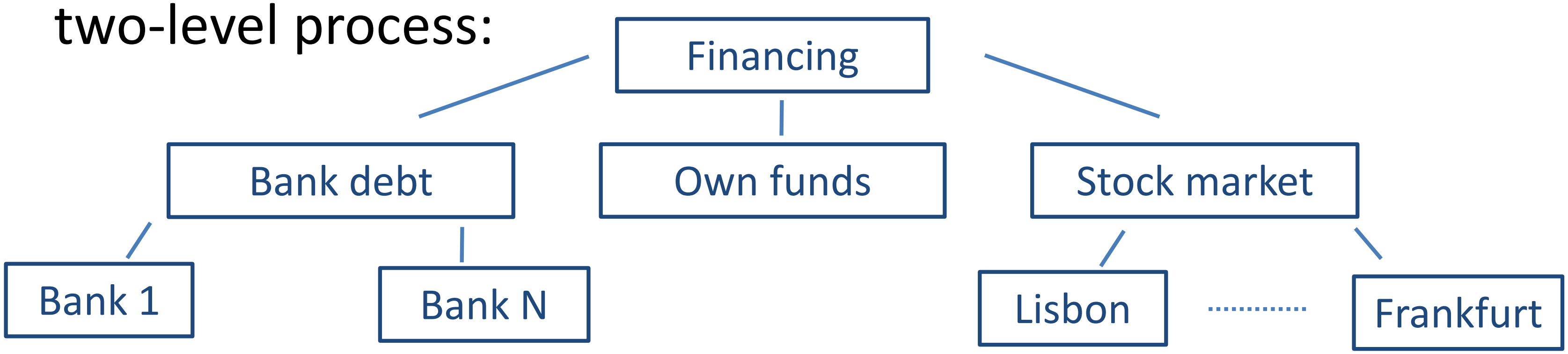
- Very complex, requiring the computation of $(M-1)$ integrals

- The version implemented in Stata assumes independent errors, which eliminates the only advantage of multinomial probit over multinomial logit

Stata
mprobit $Y X_1 \dots X_k$, baseoutcome(*0*)

Nested logit:

- Not affected by the IIA property, grouping the choices in several sets in such a way that:
  - Within each group, alternatives may be correlated
  - Between groups, alternatives are independent
- Results from a sequential decision process – example for a two-level process:



  - Level 1 – defining *J* groups,
  - Level 2 – defining M_j choices in each group

Stata

nlogit ...

Random parameters logit:

- Latent model:

$$U_{im} = x'_{im}\beta_i + u_{im}$$

- Most common assumption: $\beta_i \sim N\left(\beta, \Sigma_\beta\right)$

- Not affected by the IIA property

- If $\Sigma_\beta = 0$, it reduces to the Multinomial Logit model; hence, comparing the two models allows the IIA property to be tested

4. Models for Continuous Limited Dependent Variables

4.1. Models for Nonnegative Outcomes

4.2. Models for Fractional Responses

4.3. Models for Discrete-Continuous Responses

- Nonnegative outcomes can be:
  - Continuous: $Y \in [0, +\infty[$
    - Examples: prices, wages,…
  - Discrete (counts): $Y \in \{0,1,2,3, \dots\}$
    - Examples: patents applied for by a firm in a year, times someone is arrested in a year,...

- Linear regression models are not the most suitable option because:
  - May generate negative predictions for the dependent variable
  - At least close to the lower bound of Y, it does not make sense to assume constant partial effects

Log-linear regression model:

$$\ln(Y_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

- With this transformation, the dependent variable becomes unbounded: $Y \in]0, +\infty[\Longrightarrow \ln(Y) \in]-\infty, +\infty[$
- Assumption: $E(u_i|x) = 0$
- However, two new problems arise:
  - The log-linear model is not defined for $Y = 0$; adding a small constant value to Y or dropping zeros are not in general good solutions
  - Prediction is more interesting in the original scale, $\widehat{Y}_i$, and not in the logarithmic scale, $\widehat{\ln(Y_i)}$; the log-linear model gives the latter directly but retransforming it to the original scale requires additional assumptions and calculations and/or the application of relatively complex methods (see the next slide to understand the problem)

- Assumed model:

$$\ln(Y_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

  - Consistent estimation requires $E(u_i|x) = 0$
  - Under $E(u_i|x) = 0$:

$$E[\ln(Y_i)|x] = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}$$
$$\widehat{\ln(Y_i)} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$$

- Prediction of Y_i:
  - If $\ln(Y_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$, then:

$$Y_i = e^{\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i}$$

  and

$$E(Y_i|x) = e^{\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}} E(e^{u_i}|x)$$

  - Consistent prediction of Y_i would require assuming $E(e^{u_i}|x) = 1$; however, the assumption made, $E(u_i|x) = 0$, implies that, in general, $E(e^{u_i}|x) \neq 1$
  - Alternatively, we need to get a consistent estimate of $E(e^{u_i}|x)$, which requires additional assumptions

Exponential regression model:

$$Y = \exp(x'\beta + u)$$
$$E(Y|X) = \exp(x'\beta)$$

- Assumption: $E(e^u|x) = 1$
- Advantages:
  - $\widehat{Y}_i$ is always nonnegative
  - Predictions are obtained directly in the original scale, without requiring any retransformations
- Partial effects:

$$\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j \exp(x'\beta)$$

  - The sign of the effect is given by the sign of β_j
  - β_j can be interpreted as a semi-elasticity (see the next slide for a proof)

$$\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j \exp(x'\beta)$$

$$\Longrightarrow \Delta E(Y|X) = \beta_j E(Y|X)$$

$$\Longrightarrow \frac{\Delta E(Y|X)}{E(Y|X)} = \beta_j$$

$$\Longrightarrow 100\frac{\Delta E(Y|X)}{E(Y|X)} = 100\beta_j$$

$$\Longrightarrow \%\Delta E(Y|X) = 100\beta_j\%$$

- Assumptions and estimation methods according to the type of nonnegative outcome:
  - Continuous response:
    - Assumption: only $E(Y|X)$; estimation: QML
  - Count data - two alternatives:
    - Assumption: only $E(Y|X)$; estimation: QML
    - Assumption: $E(Y|X)$ and $Pr(Y = j|X)$; estimation: ML

- Three main distribution functions are used as basis for QML and/or ML estimation:
  - Poisson
  - Negative Binomial 1
  - Negative Binomial 2

Poisson regression model:

$$Y_i \sim Poisson(\lambda_i) \Longrightarrow Pr(Y_i = y|x_i) = \frac{e^{-\lambda_i}{\lambda_i}^y}{y!}$$

where $\lambda_i = E(Y|X) = \exp(x'\beta)$

- Estimation methods: ML (only count data) or QML, since the Poisson distribution belongs to the linear exponential family

- By definition, $E(Y|X) = Var(Y|X)$ (equidispersion), which may be a strong assumption is some empirical applications

<u>Stata</u>

ML: poisson $Y X_1 \ldots X_k$

QML: poisson $Y X_1 \ldots X_k$, robust

Negative binomial regression models:

- Two variants, both allowing for overdispersion ($\delta > 0$):
  - NEGBIN1: $Var(Y|X) = (1+\delta)E(Y|X)$ - ML estimation
  - NEGBIN2: $Var(Y|X) = [1+\delta E(Y|X)]E(Y|X)$ - it belongs to the linear exponential family, enabling estimation by both ML (only count data) and QML

Stata

NEGBIN1: nbreg $YX_1 \dots X_k$, dispersion(constant)
NEGBIN2 (ML): nbreg $YX_1 \dots X_k$, dispersion(mean)
NEGBIN2 (QML): nbreg $YX_1 \dots X_k$, dispersion(mean) robust

- Overdispersion test:

  $H_0\!: \delta = 0$ (Poisson model)

  $H_1\!: \delta \neq 0$ (Negative Binomial 1 or 2 model)

Base panel data model:

- Continuous / count data:

$$E(Y_{it}|x_{it}, \alpha_i) = \exp(\gamma_i + x_{it}'\beta) = \alpha_i \exp(x_{it}'\beta)$$

- Count data:

$$Pr(Y_{it} = y|x_{it}, \alpha_i) = \frac{e^{-\lambda_{it}} {\lambda_{it}}^{y}}{y!}$$

$$\lambda_i = E(Y_{it}|x_{it}, \alpha_i) = \alpha_i \exp(x_{it}'\beta)$$

Pooled estimator:

- Based on the cross-sectional assumption $E(Y_{it}|x_{it}) = \exp(x_{it}'\beta)$
- Produces consistent estimators only if $E(\alpha_i|x_{it}) = 1$

Stata
poisson $Y X_1 \ldots X_k$, vce(cluster *clustvar*)

Random Effects Poisson Estimator:

- Assumptions:
  - $Y_{it} \sim Poisson(\lambda_{it})$
  - $\lambda_i = E(Y_{it}|x_{it}, \alpha_i) = \alpha_i \exp(x'_{it}\beta)$
  - $\log(\alpha_i) = \gamma_i \sim Gamma(1, \eta)$

- Resulting model:
  - NEGBIN2-type model
  - Estimation method: ML
  - $E(Y_{it}|x_{it}) = \exp(x'_{it}\beta)$, which implies that the Pooled estimator is consistent under random effects of this type

Stata
xtpoisson $Y X_1 \ldots X_k$, re vce(robust)

Fixed Effects Estimators:

- Fixed effects Poisson estimator (three equivalent versions):
  - Pooled estimator with individual effects
  - Estimator conditional on $\sum_{t=1}^{T} Y_{it}$, with $\sum_{t=1}^{T} Y_{it} \neq 0$
  - Quasi mean-differenced GMM estimator (Hausman, Hall and Griliches, 1984)

- Quasi-differences GMM estimator:
  - Chamberlain (1992)
  - Wooldridge (1997)

Fixed effects Poisson estimator:

- May be derived using the three equivalent versions
- Pooled estimator with individual effects:
  - Adds individual dummies, associated to the γ_i's
  - As in linear models, β is consistently estimated even in short panels (no incidental parameters problem)
- The quasi mean-differenced GMM estimator is based on the following moment condition:

$$E\left(Y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{Y}_i \middle| x_{it}\right) = 0,$$

  where $\lambda_{it} = exp(x_{it}'\beta)$
- Requires strictly exogenous explanatory variables

Stata
xtpoisson $Y X_1 \ldots X_k$, fe vce(robust)

Quasi-differences GMM estimator :

- Chamberlain (1992):

$$E\left(\frac{\lambda_{it}}{\lambda_{i,t-1}}Y_{it} - Y_{i,t-1}\middle| x_{it}\right) = 0$$

- Wooldridge (1997):

$$E\left(\frac{Y_{it}}{\lambda_{it}} - \frac{Y_{i,t-1}}{\lambda_{i,t-1}}\middle| x_{it}\right) = 0$$

- In both cases the explanatory variables do not need to be strictly exogenous, so these estimators are particularly useful in dynamic models

Fractional outcomes:

$$Y \epsilon [0,1]$$

Base specification:

$$E(Y|X) = G(x'\beta)$$

where the $G(\cdot)$ function must respect the restriction $0 \leq G(\cdot) \leq 1$

Main models:

- Fractional regression model: assumes only $E(Y|X)$
- Beta regression model: assumes also $Pr(Y|X)$
- Transformation regression models (assume only $E(Y|X)$):
  - Linear transformation
  - Exponential transformation

Fractional regression models:

- Very similar to binary regression models
  - Main models: Logit, Probit, Cloglog
  - Partial effects calculated using the same expressions
  - Estimation also based on the Bernoulli function, but only by QML

Stata

glm $Y X_1 \ldots X_k$, family(binomial) link(logit) robust
glm $Y X_1 \ldots X_k$, family(binomial) link(probit) robust
glm $Y X_1 \ldots X_k$, family(binomial) link(cloglog) robust

Beta regression model:

- Assumes also $E(Y|X) = G(x'\beta)$, using the same functions for $G(\cdot)$
- Additional assumption: $Y_i \sim Beta$, with mean given by $G(x'\beta)$ and precision parameter ϕ
- Estimation only by ML: more efficient, less robust
- Only available when $Y \epsilon]0,1[$

Linear transformation:

$$Y_i = G(x_i'\beta + u_i)$$
$$H(Y_i) = x_i'\beta + u_i$$

- Alternative specifications:
  - Logit: $H(Y_i) = \ln \frac{Y_i}{1-Y_i}$
  - Probit: $H(Y_i) = \Phi^{-1}(Y_i)$
  - Cloglog: $H(Y_i) = \ln[-\ln(1-Y_i)]$
- Advantages:
  - Estimation: OLS
  - Easy to deal with panel data and endogenous variables
- Limitations:
  - $H(Y_i)$ is not defined for $Y_i = 0$ and $Y_i = 1$
  - Prediction in the original scale requires additional assumptions and calculations and/or the application of relatively complex methods

Example for logit:

$$Y_i = \frac{e^{x_i'\beta + u_i}}{1 + e^{x_i'\beta + u_i}}$$
$$Y_i + Y_i e^{x_i'\beta + u_i} = e^{x_i'\beta + u_i}$$
$$Y_i = e^{x_i'\beta + u_i} - Y_i e^{x_i'\beta + u_i}$$
$$Y_i = (1 - Y_i) e^{x_i'\beta + u_i}$$
$$\frac{Y_i}{1 - Y_i} = e^{x_i'\beta + u_i}$$
$$\ln \frac{Y_i}{1 - Y_i} = x_i'\beta + u_i$$

Exponential transformation:

$$Y_i = G(x_i'\beta + u_i) = G_1[\exp(x_i'\beta + u_i)]$$
$$H_1(Y_i) = \exp(x_i'\beta + u_i)$$

- Alternative specifications:
  - Logit: $H_1(Y_i) = \frac{Y_i}{1-Y_i}$
  - Cloglog: $H_1(Y_i) = -\ln(1 - Y_i)$
- Advantages:
  - Estimation: same methods as those used for nonnegative responses
  - Easy to deal with panel data and endogenous variables
- Limitations:
  - Not aplicable to the probit model
  - $H(Y_i)$ is not defined for $Y_i = 1$ (but it is for $Y_i = 0$)
  - Prediction in the original scale requires additional assumptions and calculations and/or the application of relatively complex methods

Multivariate fractional outcomes:

- $Y_{im} \in [0,1], m = 0, \ldots, M-1$
- $\sum_{m=0}^{M-1} Y_{im} = 1$

Base specification:

$$E(Y_{im}|X_i) = G_m(x'\beta)$$

- The $G_m(\cdot)$ function must respect the restrictions $0 \leq G_m(\cdot) \leq 1$ and $\sum_{m=0}^{M-1} G_m = 1$

Main models:

- Multivariate fractional regression model
- Dirichlet regression model

Multivariate fractional regression model:

- Very similar to multinomial choice models
  - Main models: Logit Multinomial, Nested Logit, Random Parameters Logit, …
  - Partial effects calculated using the same expressions
- QML estimation based on the multivariate Bernoulli function

Dirichlet regression model:

- Assumes the same specifications for $G_m(\cdot)$
- Additional assumption: $Y_i \sim Dirichlet$, with means given by $G_m(x'\beta)$ and precision parameter ϕ
- Estimation only by ML: more efficient, less robust
- Only available when $Y_{im} \in]0,1[$

Panel data - base specification:

$$E(Y_{it}|x_{it}, \alpha_i) = G(\alpha_i + x'_{it}\beta)$$

Estimators:

- Pooled estimator (requires $\alpha_i = \alpha$ for consistency)
- Pooled with individual effects (requires $T \longrightarrow \infty$ for consistency)
- Random effects (assumes $\alpha_i \sim N(0, \sigma_\alpha^2)$)
- Fixed effects (based on linear or exponential transformations)

- Tobit Model
- Two-Part Model
- Sample Selection Model

Motivation:

- Sometimes, the dependent variable has both discrete and continuous values; typically:
  - Discrete value: for many individuals, $Y_i = 0$
  - Continuous component: for the remaining individuals, Y_i may take on some positive value, which may be bounded (fractional outcome) or not (nonnegative outcome)

- Examples:
  - Expenditures on durable goods, alcohol,,...
  - Work hours

Alternative models:

- Tobit model: a single model explains all values
- Two-part model: uses two independent models for explaining separately the zeros and the positive values
- Sample selection model: uses two different, but interdependent, models for explaining the zeros and the positive values

Tobit model - specification:

- Latent model: $Y_i^* = x_i'\beta + u_i, \quad -\infty < Y_i^* < +\infty$
- Instead of Y_i^*, it is observed:

$$Y_i = \begin{cases} 0 & \text{if } Y_i^* \leq 0 \\ Y_i^* & \text{if } Y_i^* > 0 \end{cases}$$

- Assumption: $u_i \sim N(0, \sigma^2)$
  - $Pr(Y_i = 0|x_i) = Pr(Y_i^* \leq 0|x_i) = Pr(x_i'\beta + u_i \leq 0|x_i) = Pr(u_i \leq -x_i'\beta|x_i) = Pr\left(\frac{u_i}{\sigma^2} \leq -\frac{x_i'\beta}{\sigma^2}\middle| x_i\right) = \Phi\left(-\frac{x_i'\beta}{\sigma^2}\right) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma^2}\right)$
  - Hence: $f(y_i|x_i) = \begin{cases} 1 - \Phi\left(\frac{x_i'\beta}{\sigma^2}\right) & \text{if } \mathrm{Y} = 0 \\ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y_i - x_i'\beta\right)^2}{2\sigma^2}} & \text{if } Y > 0 \end{cases}$

Estimation:

- Method: ML

Stata
tobit $Y X_1 \ldots X_k$, ll(0)

- Parameters to be estimated: β and σ

- Log-likelihood function:

$$LL = \sum \left\{ (1-d_i) log \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma^2}\right) \right] + d_i log \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}} \right] \right\}$$

where $d_i = \begin{cases} 0 \text{ if } Y_i = 0 \\ 1 \text{ if } Y_i > 0 \end{cases}$

Quantities of interest:

- Conditional mean given that Y_i is positive:

$$E(Y_i|x_i, Y_i > 0) = x_i'\beta + \sigma\lambda\left(\frac{x_i'\beta}{\sigma}\right)$$

where $\lambda\left(\frac{x_i'\beta}{\sigma}\right) = \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}$ is the Mills ratio

- Probability of observing positive values for Y_i:

$$\Pr(Y_i > 0|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

- Overall conditional mean:

$$E(Y_i|x_i) = Pr(Y_i = 0|x_i)E(Y_i|x_i, Y_i = 0) + \Pr(Y_i > 0|x_i)E(Y_i|x_i, Y_i > 0)$$
$$= \Phi\left(\frac{x_i'\beta}{\sigma}\right)x_i'\beta + \sigma\phi\left(\frac{x_i'\beta}{\sigma}\right)$$

Partial effects:

- $\Delta X_j = 1 \Longrightarrow$
  - $\Delta E(Y_i|x_i, Y_i > 0) = \beta_j \left\{1 - \lambda\left(\frac{x_i'\beta}{\sigma}\right)\left[\frac{x'_i\beta}{\sigma} + \lambda\left(\frac{x_i'\beta}{\sigma}\right)\right]\right\}$
  - $\Delta Pr(Y_i > 0|x_i) = \frac{\beta_j}{\sigma}\phi\left(\frac{x_i'\beta}{\sigma}\right)$
  - $\Delta E(Y_i|x_i\ ) = \beta_j \Phi\left(\frac{x_i'\beta}{\sigma}\right)$
- The three effects have the same sign

Two-part model specification:

- First part – binary regression model:

$$Pr(d_i = 1|x_i) = G_1(x_i'\beta)$$

  - $d_i = \begin{cases} 0 \text{ se } Y_i = 0 \\ 1 \text{ se } Y_i > 0 \end{cases}$

- Second part – exponential or fractional regression model

$$E(Y_i|x_i, d_i = 1) = G_2(x_i'\theta)$$

- Overall conditional mean:

$$E(Y_i|x_i)$$
$$= Pr(Y_i = 0|x_i)E(Y_i|x_i, Y_i = 0) + \Pr(Y_i > 0|x_i)E(Y_i|x_i, Y_i > 0)$$
$$= G_1(x_i'\beta)G_2(x_i'\theta)$$

Estimation:

- Each part of the model is estimated separately:
  - In each part, use the standard methods for the type of data being analyzed
  - In the first part of the model, use the full sample
  - In the second part of the model, use the subsample for which $Y_i > 0$
  - One may use different explanatory variables in each part of the model

Partial effects:

- $\Delta Pr(d_i = 1|x_i)$
- $\Delta E(Y_i|x_i, d_i = 1)$
- $\Delta E(Y_i|x_i\ ) = \Delta Pr(d_i = 1|x_i)E(Y_i|x_i, d_i = 1) + Pr(d_i = 1|x_i)\Delta E(Y_i|x_i, d_i = 1)$

Sample selection model - latent variable:

- Y_{2i}^*: main variable
- Y_{1i}^*: variable that determines whether Y_{2i}^* is observed or not

Two equations:

- Participation equation (*e.g.* to work or not):

$$Y_{1i} = \begin{cases} 0 \text{ if } Y_{1i}^* \leq 0 \\ 1 \text{ if } Y_{1i}^* > 0 \end{cases}$$

- Outcome equation (*e.g.* how much to work):

$$Y_{2i} = \begin{cases} - \quad \text{if } Y_{1i}^* \leq 0 \\ Y_{2i}^* \text{ if } Y_{1i}^* > 0 \end{cases}$$

Latent linear models:

$$\begin{cases} Y_{1i}^* = x_{1i}'\beta_1 + u_{1i} \\ Y_{2i}^* = x_{2i}'\beta_2 + u_{2i} \end{cases}$$

Assumptions:

- The error terms of the two equations are assumed to be correlated, having a bivariate normal distribution:

$$\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \sim N\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right\}$$

- Only when $\sigma_{12} = 0$ the two equations will be independent (the selection mechanism is exogenous or ignorable):
  - In this case, the second equation may be estimated by OLS using only the observed data

Quantities of interest:

- Conditional mean of the main latent variable:

$$E(Y_{2i}^{*}|x_i) = x_{2i}'\beta_2$$

- Conditional mean of the main observed dependent variable:

$$E(Y_{2i}|x_i, Y_{1i} = 1) = x_{2i}'\beta_2 + \sigma_{12}\lambda(x_{1i}'\beta_1)$$

- Probability of observing positive values:

$$Pr(Y_{2i} > 0|x_i) = Pr(Y_{1i} = 1|x_i) = \Phi(x_{1i}'\beta_1)$$

Parameters to be estimated: $\beta, \sigma_{12}, \sigma_2$

Estimation methods:

- ML
- Heckman's two-step method

ML:

- Based on the following log-likelihood function:

$$LL = \sum \{(1 - d_i)\Pr(Y_{1i} = 0|x_{1i}) + d_i[f(Y_{1i} = 1|Y_{2i}) + f(Y_{2i})]\}$$

Stata
heckman Y_2 X_1 … X_k, select(Y_1 X_1 … X_k)

Heckman's two-step method:

- Based on $E(Y_{2i}|x_i, Y_{1i} = 1) = x'_{2i}\beta_2 + \sigma_{12}\lambda(x'_{1i}\beta_1)$
- First step: estimate the probit model $Pr(Y_{1i} = 1|x_i) = \Phi(x'_{1i}\beta_1)$ and get $\lambda\left(x'_{1i}\hat{\beta}_1\right) = \frac{\phi(x'_{1i}\hat{\beta}_1)}{\Phi(x'_{1i}\hat{\beta}_1)}$
- Second step: regress Y_{2i} on x_{2i} and $\lambda\left(x'_{1i}\hat{\beta}_1\right)$ using only individuals fully observed and OLS, and correct the variances
- t test for $\mathrm{H}_0: \sigma_{12} = 0$ (exogenous selection mechanism)
- If the same regressors are used in both steps, multicolinearity may arise; to avoid it, it is usual to exclude from x_{2i} some of the variables included in x_{1i}

Stata
heckman Y_2 X_1 ... X_k, twostep select(Y_1 X_1 ... X_k)