Ninth ARTNeT Capacity Building Workshop for Trade Research
"Trade Flows and Trade Policy Analysis"

June 2013

Bangkok, Thailand

Cosimo Beverelli and Rainer Lanz

(World Trade Organization)

Selected econometric methodologies and STATA applications

Content

a) Classical regression model
   - Linear prediction
   - Ordinary least squares estimator (OLS)
   - OLS in STATA
   - Interpretation of coefficients

b) Introduction to panel data analysis: fixed effects
   - Definition and advantages
   - Fixed effects versus random effects models
   - Fixed effects (FE) estimator
   - Alternative: Brute force OLS (dummy variable estimator)
   - FE and dummy variable estimator estimation in STATA

c) Binary dependent variable models
   i. Cross-section
   ii. Panel data
   iii. Binary dependent variable models in STATA

a) Classical regression model

- Linear prediction
- Ordinary least squares estimator (OLS)
- OLS in STATA
- Interpretation of coefficients

Linear prediction

1. Starting from an economic model and/or an economic intuition, the purpose of regression is to test a theory and/or to estimate a relationship

2. Regression analysis studies the conditional prediction of a dependent (or endogenous) variable y given a vector of regressors (or predictors or covariates) $\boldsymbol{x}$, $E[y|\boldsymbol{x}]$

3. The classical regression model is:
   - A stochastic model: $y = E[y|\boldsymbol{x}] + \varepsilon$, where ε is an error (or disturbance) term
   - A parametric model: $E[y|x] = g(\boldsymbol{x}, \beta)$, where $g(\cdot)$ is a specified function and β a vector of parameters to be estimated
   - A linear model in parameters: $g(\cdot)$ is a linear function, so: $E[y|\boldsymbol{x}] = \boldsymbol{x}'\beta$

Ordinary least squares (OLS) estimator

- With a sample of N observations ($i = 1, \dots, N$) on y and $\boldsymbol{x}$, the linear regression model is:

$$y_i = x'_i\beta + \varepsilon_i$$

where x_i is a $K \times 1$ regression vector and β is a $K \times 1$ parameter vector (the first element of x_i is a 1 for all i)

- In matrix notation, this is written as $y = X\beta + \varepsilon$

- OLS estimator of β minimizes the sum of squared errors:

$$\sum_{i=1}^{N} {\varepsilon_i}^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

which (provided that X is of full column rank K) yields:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y = \left(\sum_i x_i x'_i\right)^{-1} \left(\sum_i x_i y_i\right)$$

- This is the best linear predictor of y given $\boldsymbol{x}$ if a squared loss error function $L(e) = e^2$ is used (where $e \equiv y - \hat{y}$ is the prediction error)

OLS in STATA

- Stata's regress command runs a simple OLS regression
  - *Regress depvar indepvar1 indepvar2 …., options*

- Always use the option robust to ensure that the covariance estimator can handle heteroskedasticity of unknown form

- Usually apply the cluster option and specify an appropriate level of clustering to account for correlation within groups

- Rule of thumb: apply cluster to the most aggregated level of variables in the model
  - Example: In a model with data by city, state, and country, cluster by country

Interpretation of coefficients

- Economists are generally interested in marginal effects and elasticities

- Consider the model:

$$y = \beta x + \varepsilon$$

- $\beta = \frac{\partial y}{\partial x}$ gives the marginal effect of x on y

- If there is a dummy variable D, the model is:

$$y = \beta x + \delta D + \varepsilon$$

- $\delta = \frac{\partial y}{\partial D}$ gives the difference in y between the observations for which $D = 1$ and the observations for which $D = 0$
  - Example: if y is firm size and $D = 1$ if the firm exports (and zero otherwise), the estimated coefficient on D is the difference in size between exporters and non-exporters

Interpretation of coefficients (ct'd)

- Often, the baseline model is not a linear one, but is based on exponential mean:
$$y = \exp(\beta x)\varepsilon$$

- This implies a log-linear model of the form:
$$\ln(\mathrm{y}) = \beta x + \ln(\varepsilon)$$
  - $100 * \beta$ is the semi-elasticity of y with respect to x (percentage change in y following a marginal change in x)

- If the log-linear model contains a dummy variable:
$$\ln(\mathrm{y}) = \beta x + \delta D + \ln(\varepsilon)$$
  - The percentage change (p) in y from switching on the dummy is equal to $\exp(\hat{\delta}) - 1$
  - You can do better and estimate $\hat{p} = \exp[\hat{\delta} - \frac{1}{2} var(\hat{\delta})] - 1$, which is consistent and (almost) unbiased

Interpretation of coefficients (ct'd)

- In many applications, the estimated equation is log-log:

$$\ln(y) = \beta \ln(x) + \varepsilon$$

- β is the elasticity of y with respect to x, i.e. percentage change in y following a unit percentage increase in x

- Notice that dummies enter linearly in a log-log model, so their interpretation is the one given in the previous slide

b) Introduction to panel data analysis

- Definition and advantages
- Fixed effects (FE) versus random effects (RE) models
- Fixed effects estimator (within transformation)
- Alternative to estimate FE: Brute force OLS (dummy variables estimator)
- Fixed effects estimation in STATA (within transformation)
- Brute force OLS (dummy variables) estimation in STATA

Definition and advantages

- Panel data are repeated observations on the same cross section
  - Example: a cross-section of N firms observed over T time periods

- There are three advantages of panel data:

1. Increased precision in the estimation

2. Possibility to address omitted variable problems

3. Possibility of learning more about dynamics of individual behavior
   - Example: in a cross-section of firms, one may determine that 20% are exporting, but panel data are needed to determine whether the same 20% export each year

Fixed effects (FE) versus random effects (RE) model

- We start from a panel data model with observable variables x_{it} that vary over time t and unobserved cross-section effects α_i:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}, \qquad i = 1, \dots, N, \qquad t = 1, \dots, T$$

- The individual-specific effects α_i are random variables that capture unobserved heterogeneity
- The fixed and random effects model differ in their assumptions regarding the unobserved individual effect α_i
  - RE model: α_i are not correlated with x_{it}: Cov(x_{it}, α_i)=0
    - $\hat{\beta}_{RE}$ are efficient and consistent if assumption holds; otherwise not consistent as the error term will be correlated with the x_{it}
    - Estimation by Generalized Least Squares (GLS) (not covered as rarely used)
  - FE model: α_i are correlated with x_{it}
    - $\hat{\beta}_{FE}$ efficient if FE assumption holds; consistent under both FE and RE assumptions
    - Typically used in practice

Fixed effects estimator (within transformation)

- Take the model:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}$$

- Take the individual average over time:

$$\bar{y}_i = \alpha_i + \bar{x}'_i\beta + \bar{\varepsilon}_i$$

- Subtracting the two equations we obtain:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- OLS estimation of this equation gives the within-estimator (also called fixed effects estimator) $\hat{\beta}_{FE}$
- $\hat{\beta}_{FE}$ measures the association between individual-specific deviations of regressors from their individual-specific time averages and individual-specific deviations of the dependent variable from its individual-specific time average

Fixed effects estimator (ct'd)

- There are two potential problems for statistical inference: heteroskedasticity and autocorrelation

- Correct statistical inference must be based on panel-robust sandwich standard errors
  - Stata command: *vce(cluster id)* or *robust cluster(id)*, where id is your panel variable
  - For instance, if you observe firms over time, your id variable is the firm identifier

- You can also use panel bootstrap standard errors, because under the key assumption that observations are independent over i, the bootstrap procedure of re-sampling with replacement over i is justified
  - Stata command: *vce(bootstrap, reps(#))* where # is the number of pseudo-samples you want to use

Fixed effects estimator (ct'd)

- Applying the within-transformation seen above, we do not have to worry about the potential correlation between α_i and x'_{it}

- As long as $E(\varepsilon_{it}|x_{it}, \dots, x_{it}) = 0$ (strict exogeneity) holds, $\hat{\beta}_{FE}$ is consistent
  - Note: strict exogeneity implies that the error term has zero mean conditional on past, present and future values of the regressors

- In words, fixed effects gives consistent estimates in all cases in which we suspect that individual-specific unobserved variables are correlated with the observed ones (and this is normally the case…)

- The drawback of fixed effect estimation is that it does not allow to identify the coefficients of time-invariant regressors (because if $x_{it} = x_i$, $x_{it} - \bar{x}_i = 0$)
  - Example: it is not possible to identify the effect of foreign ownership on export values if ownership does not vary over time

Alternatives to the FE estimator: Brute force OLS (Dummy variables estimator)

- Brute force OLS does not apply the within transformation but estimates the individual-specific effects α_i by including N dummy variables in the model.

- The within transformation and the dummy variables approach result in the same estimated coefficients, i.e. $\hat{\beta}_{FE}$

- However, while $\hat{\beta}_{FE}$ are consistent with fixed T as $N \rightarrow \infty$, $\hat{\alpha}_i$ are not estimated consistently in short panels
  - The information for the estimation of $\hat{\alpha}_i$ increases as $T \rightarrow \infty$, but does not increase as $N \rightarrow \infty$

- In the case of OLS, estimation of $\hat{\alpha}_i$ does not affect the consistency of $\hat{\beta}_{FE}$. However, in many non-linear panel data models such as the probit model, estimation of the incidental parameter $\hat{\alpha}_i$ leads to inconsistent estimates of $\hat{\beta}_{FE}$ (incidental parameters problem)

Fixed effects estimation in STATA (within transformation)

- A variety of commands are available for estimating fixed effects regressions
- The most efficient method is the fixed effects regression (within estimation), *xtreg*
- Stata's *xtreg* command is purpose built for panel data regressions
- Use the *fe* option to specify fixed effects
- Make sure to set the panel dimension before using the *xtreg* command, using *xtset*
- For example:
    - *xtset countries* sets up the panel dimension as countries
    - *xtreg depvar indepvar1 indepvar2 …, fe* runs a regression with fixed effects by country
- Hint: *xtset* cannot work with string variables, so use (e.g.) *egen countries = group(country)* to convert string categories to numbers

Fixed effects (within) estimation in STATA (ct'd)

- As with regress, always specify the robust option with *xtreg*
- *xtreg, robust* will automatically correct for clustering at the level of the panel variable (firms in the previous example)
- Note that *xtreg* can only include fixed effects in one dimension. For additional dimensions, enter the dummies manually

Brute force OLS (dummy variables) estimation in STATA

- The fixed effects can enter as dummies in a standard regression
  - *Regress depvar indepvar1 indepvar2 … dum1 dum2 …., options*
  - Specify *dum** to include all dummy variables with the same stem

- Stata automatically excludes one dummy if a constant is retained in the model

- Possibilities to create dummies:
  - *Quietly tabulate country, gen(c_)*
  - Will produce c_1, c_2, etc. automatically
  - Then *regress depvar indepvar1 indepvar2 … c_*, vce(cluster …)*

- Or you can use the *i.varname* command to creates dummies
  - *xi: regress depvar indepvar1 indepvar2 … i.country, vce(cluster …)*

c) Binary dependent variable models

i. Binary dependent variable models in cross-section

ii. Binary dependent variable models with panel data

iii. Binary dependent variable models in STATA

i. Binary dependent variable models in cross-section

- Binary outcome
- Latent variable
- Linear probability model (LPM)
- Probit model
- Logit model
- Marginal effects
- Odds ratio in logit model
- Maximum likelihood (ML) estimation
- Rules of thumb

Binary outcome

- In many applications the dependent variable is not continuous but qualitative, discrete or mixed:
  - Qualitative: car ownership (Y/N)
  - Discrete: education degree (Ph.D., University degree,…, no education)
  - Mixed: hours worked per day

- Here we focus on the case of a binary dependent variable
  - Example with firm-level data: exporter status (Y/N)

Binary outcome (ct'd)

- Let y be a binary dependent variable:

$$y = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

- A regression model is formed by parametrizing the probability p to depend on a vector of explanatory variables $\boldsymbol{x}$ and a $K \times 1$ parameter vector β
- Commonly, we estimate a conditional probability:

$$p_i = \Pr[y_i = 1 | \boldsymbol{x}] = F(\boldsymbol{x}_i{}'\beta) \qquad (1)$$

where $F(\cdot)$ is a specified function

Intuition for $F(\cdot)$: latent variable

- Imagine we wanted to estimate the effect of $\boldsymbol{x}$ on a continuous variable y^*
- The "index function" model we would like to estimate is:

$$y_i{}^* = \boldsymbol{x}_i{}'\beta - \varepsilon_i$$

- However, we do not observe y^* but only the binary variable y

$$y = \begin{cases} 1 & if\ y^* > 0 \\ 0 & otherwise \end{cases}$$

Intuition for $F(\cdot)$: latent variable (ct'd)

- There are two ways of interpreting ${y_i}^*$:

1. Utility interpretation: ${y_i}^*$ is the additional utility that individual i would get by choosing $y_i = 1$ rather than $y_i = 0$

2. Threshold interpretation: ε_i is a threshold such that if $\boldsymbol{x}_i{}'\beta > \varepsilon_i$, then $y_i = 1$

- The parametrization of p_i is:

$$\begin{aligned} p_i &= \Pr[y = 1|\boldsymbol{x}] = \Pr[y^* > 0|x] = \Pr[\,\boldsymbol{x}'\beta - \varepsilon > 0|x] \\ &= \Pr[\varepsilon < \boldsymbol{x}'\beta] = F[\boldsymbol{x}'\beta] \end{aligned}$$

where $F(\cdot)$ is the CDF of ε

Linear probability model (LPM)

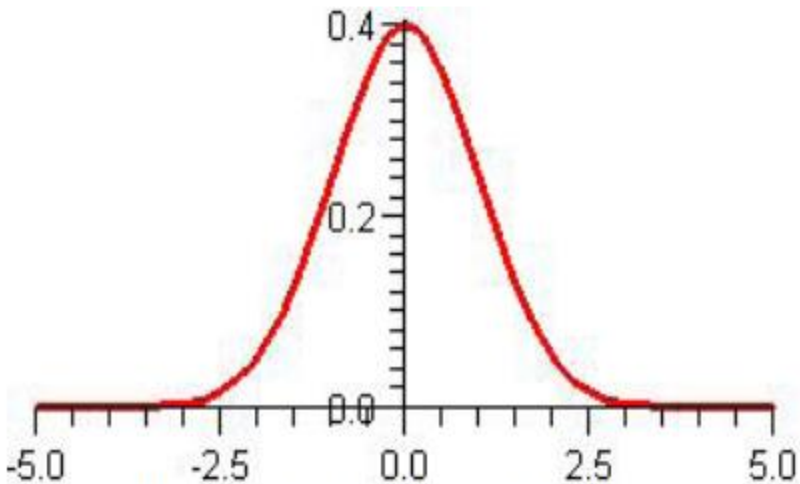
- The LPM does not use a CDF, but rather a linear function for $F(\cdot)$
- Therefore, equation (1) becomes:

$$p_i = \Pr[y_i = 1|\boldsymbol{x}] = \boldsymbol{x}_i{}'\beta$$

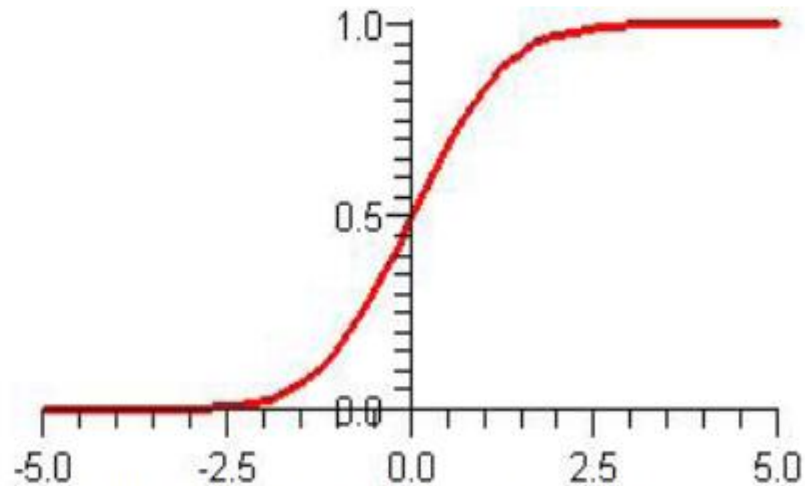
- The model is estimated by OLS with error term ε_i
- From basic probability theory, it should be the case that $0 \le p_i \le 1$
- This is not necessarily the case in the LPM, because $F(\cdot)$ in not a CDF (which is bounded between 0 and 1)
  - Therefore, one could estimate predicted probabilities $\hat{p}_i = \boldsymbol{x}_i{}'\widehat{\beta}$ that are negative or exceed 1
- Moreover, $V(\varepsilon_i) = \boldsymbol{x}_i{}'\beta(1 - \boldsymbol{x}_i{}'\beta)$ depends on $\boldsymbol{x}_i$
  - Therefore, there is heteroskedasticity (standard errors need to be robust)
- However, LPM provides a good guide to which variables are statistically significant

Probit model

- The probit model arises if $F(\cdot)$ is the CDF of the normal distribution, $\Phi(\cdot)$
- So $\Phi(\boldsymbol{x}'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz$, where $\phi(\cdot) \equiv \Phi'(\cdot)$ is the normal pdf



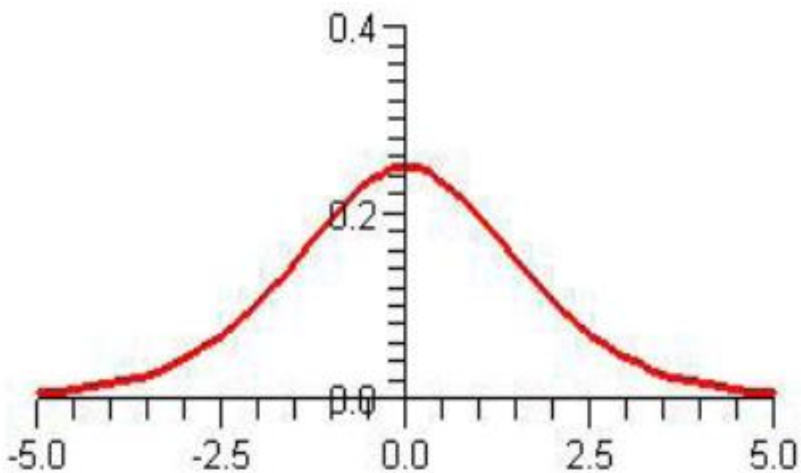
PDF of the Standard Normal Distribution



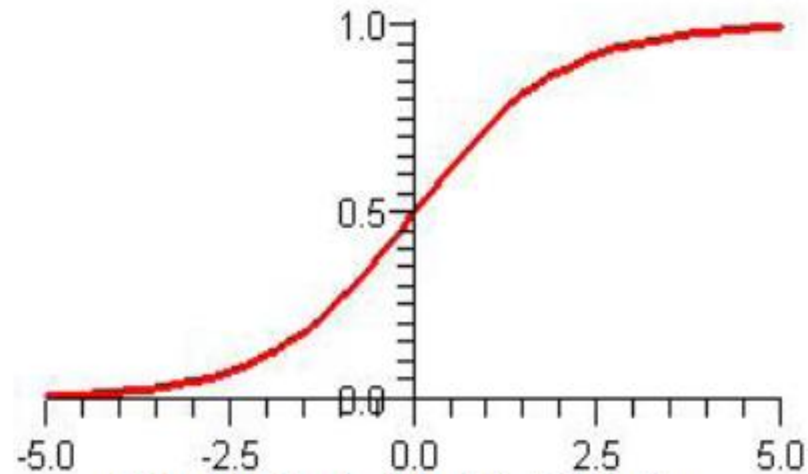
CDF of the Standard Normal Distribution

Logit model

- The logit model arises if $F(\cdot)$ is the CDF of the logistic distribution, $\Lambda(\cdot)$
- So $\Lambda(\boldsymbol{x}'\beta) = \frac{e^{x'\beta}}{1-e^{x'\beta}}$



PDF of the Standard Logistic Distribution



CDF of the Standard Logistic Distribution

Marginal effects

- For the model $p_i = \Pr[y_i = 1|\boldsymbol{x}] = F(\boldsymbol{x}_i{}'\beta) - \varepsilon_i$, the interest lies in estimating the marginal effect of the j'th regressor on p_i:

$$\frac{\partial p_i}{\partial x_{ij}} = F'(\boldsymbol{x}_i{}'\beta)\beta_j$$

- In the LPM model, $\frac{\partial p_i}{\partial x_{ij}} = \beta_j$

- In the probit model, $\frac{\partial p_i}{\partial x_{ij}} = \phi(\boldsymbol{x}_i{}'\beta)\beta_j$

- In the logit model, $\frac{\partial p_i}{\partial x_{ij}} = \Lambda(\boldsymbol{x}'\beta)[1 - \Lambda(\boldsymbol{x}_i{}'\beta)]\beta_j$

Odds ratio in logit model

- The odds ratio $\mathrm{OR} \equiv p/(1-p)$ is the probability that $y=1$ relative to the probability that $y=0$
- An odds ratio of 2 indicates, for instance that the probability that $y=1$ is twice the probability that $y=0$
- For the logit model:

$$p = e^{\boldsymbol{x}'\beta}/(1+e^{\boldsymbol{x}'\beta})$$
$$\mathrm{OR} = p/(1-p) = e^{\boldsymbol{x}'\beta}$$
$$\ln(OR) = \boldsymbol{x}'\beta$$

(the log-odds ratio is linear in the regressors)

- β_j is a semi-elasticity
- If $\beta_j = 0.1$, a one unit increase in regressor j increases the odds ratio by a multiple 0.1
- See also here

Maximum likelihood (ML) estimation

- Since y_i is Bernoulli distributed ($y_i = 0, 1$), the density (pmf) is:

$$f(y_i|x_i) = {p_i}^{y_i}(1 - p_i)^{1-y_i}$$

Where $p_i = F(\boldsymbol{x}_i{}'\beta)$

- Given independence over i's, the log-likelihood is:

$$\mathcal{L}_N(\beta) = \sum_{i=1}^{N} \{y_i \ln F(\boldsymbol{x}_i{}'\beta) + (1 - y_i) \ln(1 - F(\boldsymbol{x}_i{}'\beta))\}$$

- There is no explicit solution for $\hat{\beta}_{MLE}$, but if the log-likelihood is concave (as in probit and logit) the iterative procedure usually converges quickly
- There is no advantage in using the robust sandwich form of the VCV matrix unless $F(\cdot)$ is mis-specified
- If there is cluster sampling, standard errors should be clustered

Rules of thumb

- The different models yield different estimates $\hat{\beta}$
- This is just an artifact of using different formulas for the probabilities
- It is meaningful to compare the marginal effects, not the coefficients
- At any event, the following rules of thumb apply:

$$\hat{\beta}_{Logit} \cong 1.6\ \hat{\beta}_{Probit}$$

$$(\text{or } \hat{\beta}_{Logit} \cong (\frac{\pi}{\sqrt{3}})\ \hat{\beta}_{Probit})$$

- The differences between probit and logit are negligible if the interest lies in the marginal effects averaged over the sample

ii. Binary dependent variable models with panel data

- Individual-specific effects binary models
- Fixed effects logit

Individual-specific effects binary models

- With panel data (each individual i is observed t times), the natural extension of the cross-section binary models is:

$$p_{it} = \Pr[y_{it} = 1 | x_{it}, \beta, \alpha_i] = \begin{cases} F(\alpha_i + \boldsymbol{x}'_{it}\beta) & in\ general \\ \Lambda(\alpha_i + \boldsymbol{x}'_{it}\beta) & for\ Logit\ model \\ \Phi(\alpha_i + \boldsymbol{x}'_{it}\beta) & for\ Probit\ model \end{cases}$$

- Random effects estimation assumes that $\alpha_i \sim N(0, \sigma^2{}_\alpha)$

Individual-specific effects binary models (ct'd)

- Fixed effect estimation is not possible for the probit model because there is an incidental parameters problem
  - β are common parameters, i.e. they are common to all observations
  - α_i are incidental parameters, their estimation depends on fixed T observations
  - Estimating α_i (N of them) along with β leads to inconsistent estimators of both α_i and β; the shorter T the larger the inconsistency
  - Unconditional fixed-effects probit models may be fit with the "probit" command with indicator variables for the panels. However, unconditional fixed-effects estimates are biased

- However, fixed effects estimation is possible with logit, using a conditional MLE that uses a conditional density (which describes a subset of the sample, namely individuals that "change state")

Fixed effects logit

- A conditional ML can be constructed conditioning on $\sum_t y_{it} = c$, where $0 < c < T$
- The functional form of $\Lambda(\cdot)$ allows to eliminate the individual effects and to obtain consistent estimates of β

- Notice that it is not possible to condition on $\sum_t y_{it} = 0$ or on $\sum_t y_{it} = T$
- Observations for which $\sum_t y_{it} = 0$ or $\sum_t y_{it} = T$ are dropped from the likelihood function
- That is, only the individuals that "change state" at least once are included in the likelihood function

Example

- T = 3
- We can condition on $\sum_t y_{it} = 1$ (possible sequences $\{0,0,1\}$, $\{0,1,0\}$ and $\{1,0,0\}$ or on $\sum_t y_{it} = 2$ (possible sequences $\{0,1,1\}$, $\{1,0,1\}$ and $\{1,1,0\}$)
- All individuals with sequences $\{0,0,0\}$ and $\{1,1,1\}$ are not considered

iii. Binary dependent variable models in Stata

- Limited dependent variable models in cross section
- Panel data applications

Binary dependent variable models in cross section

- Probit
  - *Probit depvar indepvar1 indepvar2 …, options*

- Logit – two commands providing identical results:
  - *Logit depvar indepvar1 indepvar2 …, options*
  - *Logistic depvar indepvar1 indepvar2 …, options*
  - For *logit*, chose option “*or”* to show odds ratios instead of coefficients. For *logistic*, enter “*logit*” after the estimation to get the coefficients

- Generally speaking, results from the probit and logit models are quite close. Except in special cases, there is no general rule to prefer one over the other

Panel data applications

- Probit and logit can both be estimated with random effects:
- To obtain probit and logit results with random effects by “id”:
  - *xtset id*
  - *xtprobit depvar indepvar1 indepvar2 …, re*
  - *xtlogit depvar indepvar1 indepvar2 …, re*

- Logit models can be consistently estimated with fixed effects, and should be preferred to probit in panel data settings
- To obtain logit results with fixed effects by “id”:
  - *xtset id*
  - *xtlogit depvar indepvar1 indepvar2 …, fe*

- The “conditional logit” (clogit) estimation should be preferred, however, because it allows for clustered-robust standard errors