ADVANCED ECONOMETRICS I

Theory (2/3)

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

2. Nonlinear Regression Analysis

2. Nonlinear Regression Analysis

Motivation:

- Often, the dependent variable is discrete and/or bounded, in which case linear regression models cannot describe it appropriately

- Some continuous, bounded dependent variables may be transformed in such a way that linear regression models can still be used for their analysis; but in some cases such transformations are not available

2. Nonlinear Regression Analysis

Quantities of interest:

- Linear models:
  - $E(Y|X)$

- Nonlinear models:
  - $E(Y|X)$
  - If using a probabilistic model: $Pr(Y|X)$
  - In some models, there may be also interest on variants of the previous quantities:
    - Example: when modelling a nonnegative outcome, $Y \geq 0$, with lots of zeros, it may be interesting to estimate also:
      - » $Pr(Y = 0|X)$
      - » $E(Y|X, Y > 0)$

2. Nonlinear Regression Analysis

Partial Effects:

- Linear models:
  - Model: $E(Y|X) = X\beta$
  - Effects: $\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j$

- Nonlinear models:
  - Model:
    - $E(Y|X) = G(X\beta)$
    - $Pr(Y|X) = F(X\beta)$
  - Effects: $\Delta X_j = 1 \Longrightarrow$
    - $\Delta E(Y|X) = \frac{\partial E(Y|X)}{\partial X_j} = \frac{\partial G(X\beta)}{\partial X_j} = \beta_j \frac{\partial G(X\beta)}{\partial X\beta} = \beta_j g(x_i'\beta)$
    - $\Delta Pr(Y|X) = \frac{\partial Pr(Y|X)}{\partial X_j} = \frac{\partial F(X\beta)}{\partial X_j} = \beta_j \frac{\partial F(X\beta)}{\partial X\beta} = \beta_j f(x_i'\beta)$

2. Nonlinear Regression Analysis

- Partial effects may be compared across different models, but the values of β cannot
- However, because $\frac{\partial G(X\beta)}{\partial X\beta} > 0$ and $\frac{\partial F(X\beta)}{\partial X\beta} > 0$:
  - The sign of the partial effect is given by the sign of β_j
  - Testing the statistical significance of the partial effect is equivalent to test for $H_0: \beta_j = 0$
- To calculate the magnitude of the partial effects, there are three main alternatives:
  - Calculate the partial effects for each individual in the sample and then obtain the mean of those effects
  - Replace x by its sample means
  - Replace x by specific values

<u>Stata</u>
(after estimating the model)
$\text{margins, dydx}(varlist)\ \text{at}(\ldots)$
$\text{margins, dydx}(varlist)\ \text{atmeans}$
$\text{margins, dydx}(varlist)$

Estimation:

- Most common estimation methods:
  - Maximum Likelihood (ML): more efficient
  - Quasi-Maximum Likelihood (QML): more robust
- In both cases it is necessary to specify:
  - The G function in $E(Y|X) = G(X\beta)$
  - The F function in $Pr(Y|X) = F(X\beta)$
- Main assumptions:
  - ML:
    - Correct specification of both G and F
  - QML:
    - Correct specification of G
    - F does not need to be correctly specified but needs to belong to the linear exponential family (e.g. Normal, Bernoulli, Poisson, Exponencial, Gama, etc.)

ML / QML estimation - Statistics:

- Distribution function - $F(y)$: gives the probability of the random variable Y taking a value less than or equal to y:

$$F(y) = Pr(Y \leq y)$$

- Density function - $f(y)$:
  - Derivative of the distribution function:

$$f(y) = \frac{\partial F(y)}{\partial y} \qquad F(y) = \int_{-\infty}^{y} f(y) dY$$

  - In the continuous case, describes the *relative* likelihood for the random variable Y being equal to y (not the *absolute* likelihood)
  - In the discrete case gives the probability of the random variable Y being equal to y

- Likelihood function:
  - In individual terms, it is the same as the density function
  - Usually, it is calculated for the full sample, giving the *likelihood* of observing that sample under the assumption that the density function $f(y)$ describes appropriately the population behaviour
  - Assuming independence across individuals and the same distribution for all of them, it is calculated as:

$$L(y) = \prod_{i=1}^{N} f(y_i), \qquad\qquad 0 \leq L(y) \leq 1$$

- Usually:
  - $F(y)$, $f(y)$ and $L(y)$ depend also on 1 or 2 parameters
  - One of the parameters represents $E(y)$

Most popular density functions:

Y	Function	$f(y)$
$]-\infty,+\infty[$	Normal (μ,σ^2)	$\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$
$]0,+\infty[$	Exponential (μ)	$\frac{1}{\mu}e^{-\frac{y}{\mu}}$
$]0,1[$	Beta (μ,δ)	$\frac{\Gamma(\delta)}{\Gamma(\mu\delta)\Gamma[(1-\mu)\delta]}y^{\mu\delta-1}(1-y)^{(1-\mu)\delta-1}$
$\{0,1\}$	Bernoulli (μ)	$\mu^y(1-\mu)^{1-y}$
$\{0,1,2,\dots\}$	Poisson (μ)	$\frac{\mu^y e^{-\mu}}{y!}$

In all cases: $E(y)=\mu$

Econometrics:

- All the analysis is conditional on a set of explanatory variables
- The parameter μ ($= E(y)$ in Statistics) is replaced by the function assumed for $E(y|X)$, for example $X\beta$ (linear regression model)
- It is assumed that the likelihood function is known up to the set of parameters β (and, in case the original function has 2 parameters, the other parameter)
- Density function to be considered: $f(y_i|x_i;\beta)$

Estimation:

- Given that:
  - The density function $f(\cdot)$ is known, except for β
  - The probability that the sample values were in fact generated by the chosen density $f(\cdot)$ is measured by the likelihood function
- Then:
  - We should choose for β the value that maximizes $L(y_i|x_i;\beta)$
  - Optimization problem:

$$\max_\beta L(y|X;\beta) = \prod_{i=1}^{N} f(y_i|x_i;\beta)$$

  - Actually, it is more common to maximize $LL(X\beta) = \ln[L(y|X;\beta)]$:

$$\max_\beta LL(X\beta) = \sum_{i=1}^{N} \ln[f(y_i|x_i;\beta)]$$

    – It is easier to maximize
    – It produces the same estimates for β

Properties of ML / QML Estimators:

- Asymptotic properties of ML estimators:
  - Consistency
  - Efficiency
  - Normality
- Asymptotic properties of QML estimators:
  - Consistency
  - Normality
  - Efficiency is lost; variance calculated in a robust way
  - Not possible to predict $Pr(Y|X)$ and associated partial effects
- Finite sample properties for both estimators:
  - Unknown

Alternative forms for estimating parameter variances:

- Standard / Efficient $\longrightarrow$ only available for ML
- Robust $\longrightarrow$ only makes sense for QML
- Cluster-robust $\longrightarrow$ panel data
- Bootstrap

Classical tests:

- Likelihood Ratio (LR) $\longrightarrow$ only available for ML
- Wald
- Score/LM

Test for the joint significance of a set of parameters:

- Competing models:
  - Restricted (smaller) model, based on $L_R(\beta_0 + \beta_1 x_1 + \cdots + \beta_g x_g)$
  - Full (larger) model , based on $L_F(\beta_0 + \beta_1 x_1 + \cdots + \beta_g x_g + \beta_{g+1} x_{g+1} + \cdots + \beta_k x_k)$

- Hypotheses:

H_0: $\beta_{g+1} = \cdots = \beta_k = 0$ (restricted model)

H_1: No H_0 (full model)

- LR test:

$$LR = 2[LL_F(X\beta_F) - LL_R(X\beta_R)] \sim \chi^2_{k-g}$$

  - Available in most econometric packages
  - Easy calculation
  - Both the competing models need to be estimated

Stata
(estimate one model)
estimates store *Model1*
(estimate the other model)
estimates store *Model2*
lrtest *Model1 Model2*

- Wald test:

$$W = \hat{\beta}'_D\left[\mathrm{Var}\left(\hat{\beta}_D\right)\right]^{-1}\hat{\beta}_D \sim \chi^2_{k-g}$$

where $\hat{\beta}_D = \left(\hat{\beta}_{g+1}, \dots, \hat{\beta}_k\right)$ is estimated based on $LL_F(X\beta_F)$

  - When H_0: $\beta_j = 0$, W simplifies to:

$$t = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim \mathcal{N}(0,1)$$

  - Available in most econometric packages
  - Only the full model needs to be estimated

Stata
(after estimating the full model)
test $X_{g+1} \cdots X_k$

- Score/LM test:

$$\text{Score} = \frac{\partial LL_F\left(X\hat{\beta}_M\right)}{\partial \beta}\left[Var_F\left(\hat{\beta}_M\right)\right]^{-1}\frac{\partial LL_F\left(X\hat{\beta}_M\right)}{\partial \beta} \sim \chi^2_{k-g}$$

where $\hat{\beta}_M = \left(\hat{\beta}_0, \dots, \hat{\beta}_g, 0, \dots 0\right)$, with $\left(\hat{\beta}_0, \dots, \hat{\beta}_g\right)$ estimated based on $LL_R(X\beta_R)$

  - Only the restricted model needs to be estimated, which may be an advantage when the full model is complex and hard to estimate
  - Rarely available in econometric packages, requiring programming

Specification tests:

- For $E(Y|X)$:
  - RESET test
  - Chow test

- For $Pr(Y|X)$:
  - Information Matrix text, usually very hard to implement
  - More common: tests designed specifically to particular models

RESET test:

- Implementation:
  - Estimate the original model:
$$Pr(Y|X) = F(\beta_0 + \beta_1 x_1 + \cdots \beta_k x_k)$$
  - Generate the variables $\left(X\hat{\beta}\right)^2, \left(X\hat{\beta}\right)^3, \left(X\hat{\beta}\right)^4, \ldots$
  - Add the generated variables to the original model and estimate the following auxiliary model:
$$Pr(Y|X) = F\left[\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma_1\left(X\hat{\beta}\right)^2 + \gamma_2\left(X\hat{\beta}\right)^3 + \gamma_3\left(X\hat{\beta}\right)^4 + \cdots\right]$$
  - Apply a LR / Wald test for the significance of the added variables:

    $H_0: \gamma_1 = \gamma_2 = \gamma_3 = \cdots = 0$ (suitable model functional form)

    H_1: No H_0 (unsuitable model functional form)

Chow Test for Structural Breaks:

- Context:
  - Two groups of individuals / firms / ...: G_A, G_B
  - It is suspected that the behaviour of the two groups in which regards the dependent variable may have different determinants
- Implementation:
  - Generate the dummy variable $D = \begin{cases} 1 & \text{if the individual belongs to } G_A \\ 0 & \text{if the individual belongs to } G_B \end{cases}$
  - Estimate the original model 'duplicated':

$$Pr(Y|X) = F(\theta_0 + \theta_1 X_1 + \cdots + \theta_k X_k + \gamma_0 D + \gamma_1 D X_1 + \cdots + \gamma_k D X_k)$$

  - Apply a LR / Wald test for the significance of the variables where D is present:

    $H_0: \gamma_0 = \cdots = \gamma_k = 0$ (no structural break)

    H_1: Não H_0 (with a structural break)

Base nonlinear model for panel data:

- Individual effects model:

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

- Unlike the linear case:
  - Assuming $E(\alpha_i|x_{it}) = 0$ is not enough to get consistent estimators
  - In general, methods based on subtracting time averages or first-diferences do not eliminate fixed effects
  - Inconsistent estimation of α_i leads to inconsistent estimation of β (incidental parameters problem)

Main estimators:

- Pooled estimator:
  - Based on the estimation of the model $E(Y_{it}|x_{it}) = G(x_{it}'\beta)$ and $Pr(Y_{it}|x_{it}) = F(x_{it}'\beta)$, being consistent only under the assumption of no individual effects
  - Even with random effects this estimator will be, in general, inconsistent

- Pooled estimator with individual effects:
  - Adds dummies for each individual, allowing estimation of the α_i's
  - Consistent only if $T \to \infty$

- Fixed effects estimator:
  - Assumes $E(\alpha_i|x_{it}) \neq 0$
  - Long panels
    - Use the pooled estimator with individual effects
  - Short panels:
    - In a few cases:
      - » It is possible to drop the $\alpha_i's$ from the model to be estimated using methods defined on a case-by-case basis (may also be used with long panels)
      - » In general, prediction and quantification of partial effects are not possible
    - In most cases, no fixed effects estimator is available

- Random effects estimator:
  - Most popular panel data estimator for probabilistic models
  - It is necessary to:
    - Correctly specify $f(y_i|x_i, \alpha_i; \beta)$
    - Assume that α_i follows some distribution $f(\alpha_i; \eta)$
  - Density function for maximum likelihood estimation:

$$f(y_i|x_i; \beta, \eta) = \int f(y_i|x_i, \alpha_i; \beta) f(\alpha_i; \eta) d\alpha_i$$

    - In general, this expression cannot be simplified
    - Because of the integral, it requires numerical methods
  - QML estimation not available
  - In general, prediction and quantification of partial effects are not possible

3. Discrete Choice Models

3.1. Models for Binary Choices

3.2. Models for Ordered Choices

3.3. Models for Multinomial Choices

3. Discrete Choice Models

Models for:

- Binary choices:
  - $Y \in \{0,1\}$
  - Ex.: be (or not) successful in a mortgage application

- Multinomial choices
  - $Y \in \{0,1,\dots,M-1\}$
  - Ex.: choosing a brand

- Ordered choices
  - $Y \in \{0,1,\dots,M-1\}$
  - Ex.: firms getting a specific investment rating

3. Discrete Choice Models

Common structure:

- M choices

- Aim - explaining the probability of observing $Y_i = y_i$ given $X_i = x_i$:

$$Pr(Y_i = y_i | X_i = x_i) = F(x_i'\beta)$$

- Since $\sum_y Pr(Y_i = y_i | x_i) = 1$:
  - Only $M - 1$ choices are modelled, the probability of the other being obtained by difference
  - The sum of the partial effects has to be null, $\sum_y \Delta Pr(Y_i = y_i | x_i) = 0$, with one of them being obtained by difference

Binary choices:

- Dependent variable only takes on the values 0 and 1

- Bernoulli density function:

$$f(y_i|x_i) = \mu_i^{y_i}(1-\mu_i)^{1-y_i}$$
$$\mu_i = E(Y_i|x_i) = G(x_i'\beta),$$

  where $0 < G(\cdot) < 1$

- Note that $E(Y_i|x_i) = Pr(Y_i = 1|x_i)$, since:

$$E(Y_i|x_i) = 1 \times Pr(Y_i = 1|x_i) + 0 \times Pr(Y_i = 0|x_i)$$
$$= Pr(Y_i = 1|x_i);$$

  Therefore, one may choose for $G(\cdot)$ a distribution function, which, by definition, is bounded by 0 and 1

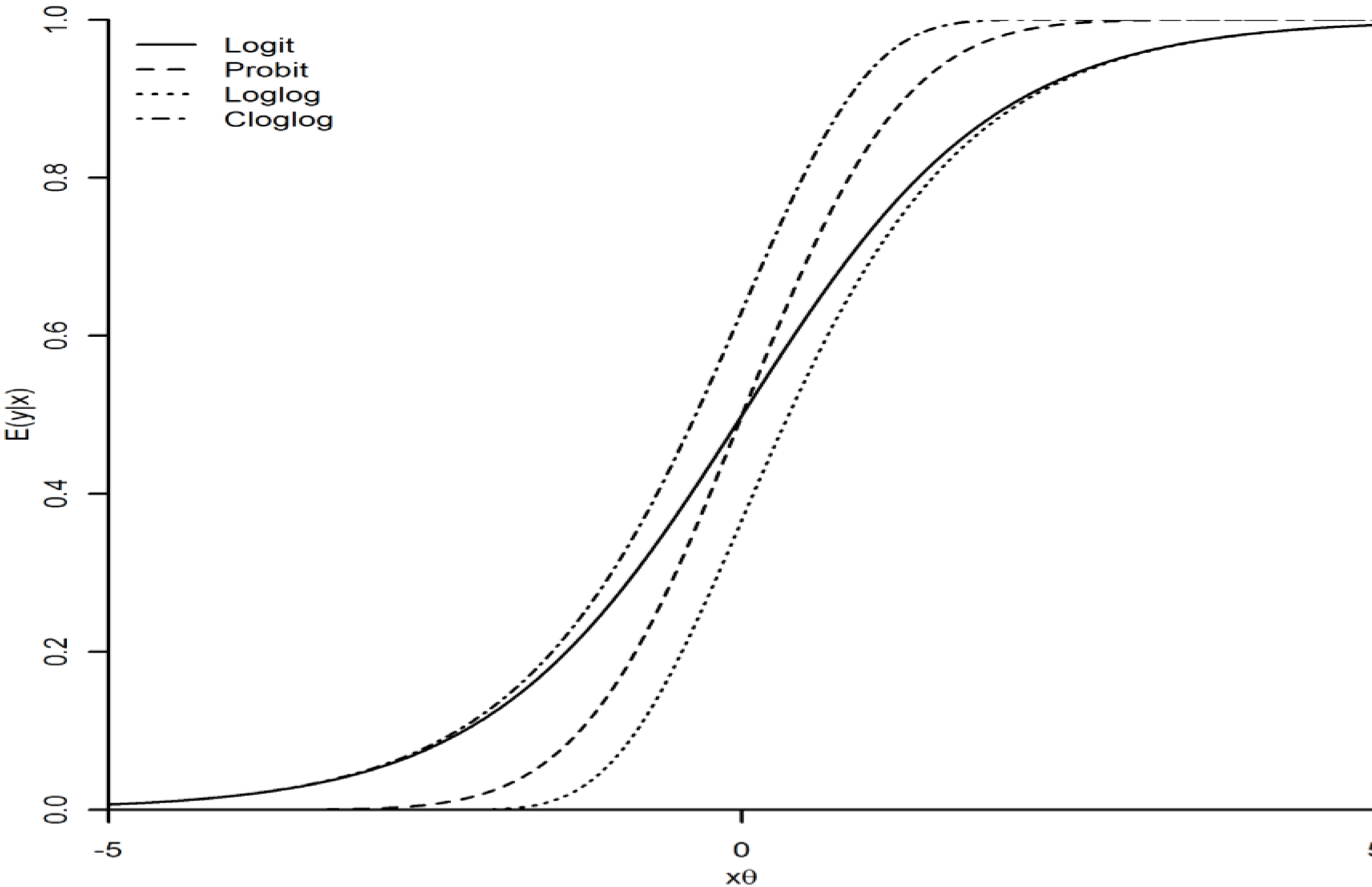
Estimation:

- QML not possible:
  - The correct specification of $E(Y_i|X_i)$ implies automatically the correct specification of $Pr(Y_i = 1|X_i)$
- Estimation by ML based on:

$$LL = \sum_{i=1}^{N} \{y_i \ln[G(x_i'\beta)] + (1 - y_i)\ln[1 - G(x_i'\beta)]\}$$

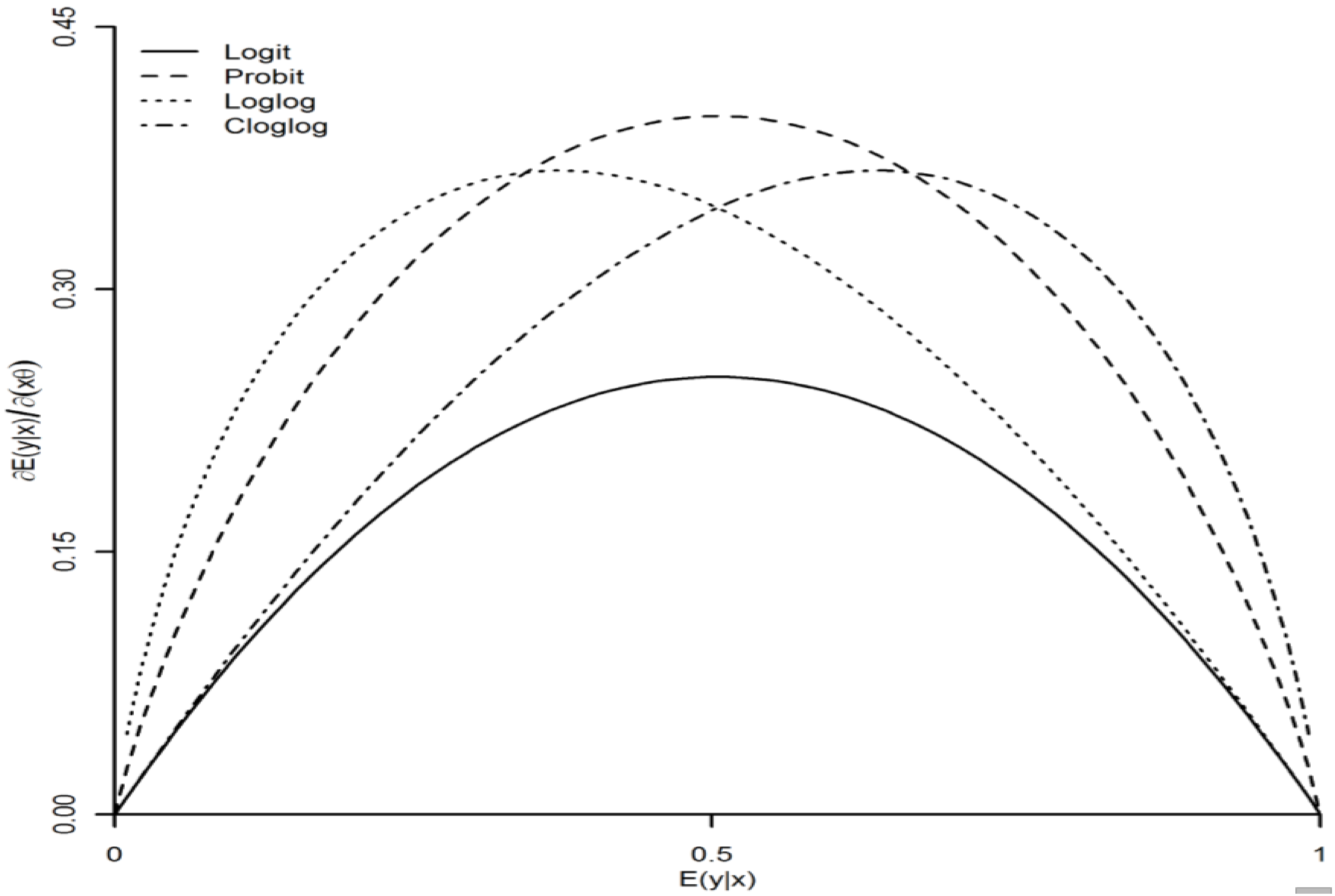
- According to the specification of G, different the resultant model – examples:
  - Probit: $G(x_i'\beta) = \Phi(x_i'\beta) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(x_i'\beta\right)^2}{2}} dx\beta$
  - Logit: $G(x_i'\beta) = \Lambda(x_i'\beta) = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$
  - Cloglog: $G(x_i'\beta) = 1 - e^{-e^{x_i'\beta}}$

Stata
logit $YX_1 \dots X_k$
probit $YX_1 \dots X_k$
cloglog $YX_1 \dots X_k$


Logit
Probit
Loglog
Cloglog
E(y|x)
xθ
0.0
0.2
0.4
0.6
0.8
1.0
-5
0
5

Partial effects:

- $\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \Delta Pr(Y = 1|X) = \frac{\partial G(X\beta)}{\partial X_j} = \beta_j \frac{\partial G(X\beta)}{\partial (X\beta)} = \beta_j g(x_i'\beta)$, with $g(x_i'\beta)$ given by:
  - Logit: $g(x_i'\beta) = \lambda(x_i'\beta) = \Lambda(x_i'\beta)[1 - \Lambda(x_i'\beta)]$
  - Probit: $g(x_i'\beta) = \phi(x_i'\beta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i'\beta)^2}{2}}$
  - Cloglog: $g(x_i'\beta) = [1 - G(x_i'\beta)]e^{x_i'\beta}$


Logit
Probit
Loglog
Cloglog
0.45
0.30
0.15
0.00
∂E(y|x)/∂(xθ)
0
0.5
1
E(y|x)

Selection criteria:

- To select the most suitable model, in addition to the RESET test, it is common to calculate the percentage of correct predictions of each model:

	$Y_i = 1$	$Y_i = 0$	**Total**
$\hat{Y}_i = 1$	n_{11}		
$\hat{Y}_i = 0$		n_{00}	
Total	n_1	n_0	n

- $\hat{Y}_i = \begin{cases} 1 \text{ if } \widehat{Pr(Y_i = 1|x_i)} \geq 0.5 \\ 0 \text{ if } \widehat{Pr(Y_i = 1|x_i)} < 0.5 \end{cases}$
- % correct predictions: $(n_{11} + n_{00})/n$
- % 1's correctly predicted: n_{11}/n_1
- % 0's correctly predicted: n_{00}/n_0

Stata
(after estimating the model)
estat classification

Alternative motivation:

- In Economics, using the utility concept to explain the optimal choices of agents is very common
- The satisfaction experienced by the consumer of a good cannot be measured accurately; however, the decision to buy or not the good can be observed
- Strategy:
  - Linear regression model to explain the difference in utilities (Y_i^*) of the two goods, using as dependent variable a continuous latent variable $Y_i^* = x_i'\beta + u_i$
  - Binary regression model to explain the probability of choosing a good:
    - Instead of Y_i^*, one observes $Y_i = \begin{cases} 0 \text{ se } Y_i^* \leq 0 \\ 1 \text{ se } Y_i^* > 0 \end{cases}$
    - Model: $Pr(Y_i = 1|x_i) = Pr(Y_i^* > 0|x_i) = Pr(x_i'\beta + u_i > 0|x_i) = Pr(u_i > -x_i'\beta|x_i) = Pr(u_i < x_i'\beta|x_i) = G(x_i'\beta)$

Panel data:

- Base model – individual effects model:

$$Pr(Y_{it} = 1|x_{it}, \alpha_i) = E(Y_{it}|x_{it}) = G(\alpha_i + x_{it}'\beta)$$

- Estimators:
  - Pooled (omits α_i; consistent only if $\alpha_i = \alpha$)
  - Pooled with individual effects (consistent only if $T \longrightarrow \infty$)
  - Random effects (assumes $\alpha_i \sim N(0, \sigma_\alpha^2)$)
  - Fixed effects logit

Stata
Pooled: same commands as for cross-sectional data
Random effects: xtprobit, xtlogit, xtcloglog

Fixed effects logit model:

- It can be shown that the α_i's may be eliminated from a logit model if the analysis is conditional on individuals for whom $\sum_{t=1}^{T} Y_{it} \neq 0$ and $\sum_{t=1}^{T} Y_{it} \neq T$:
  - All individuals whom display the value of 1 for the dependent variable in all time periods are dropped from the sample
  - The same occurs with individuals displaying always the value of 0 for the dependent variable
  - Only individuals that change their states at least once over time are relevant for estimation
- This method only works for the logit model

Stata
xtlogit $Y X_1 \ldots X_k$, fe