Models for Fractional Responses

Conditional Mean and Beta Regression Models

Transformation Regression Models

Multivariate Fractional Responses

Panel Data Models

Endogeneity

Models for Fractional Responses

Fractional outcomes:

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

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

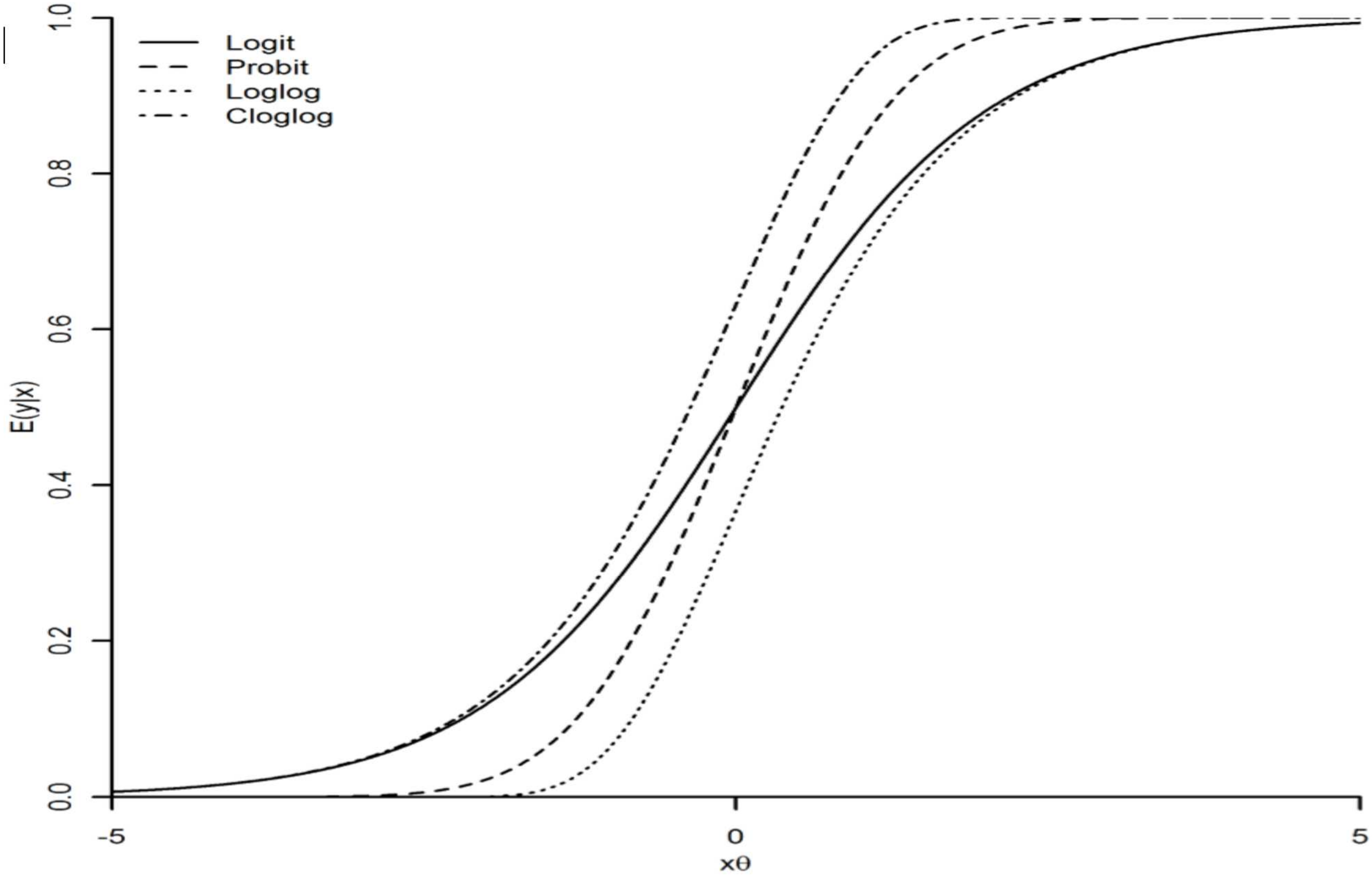
Fractional regression models:

- Estimation by QML based on:

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

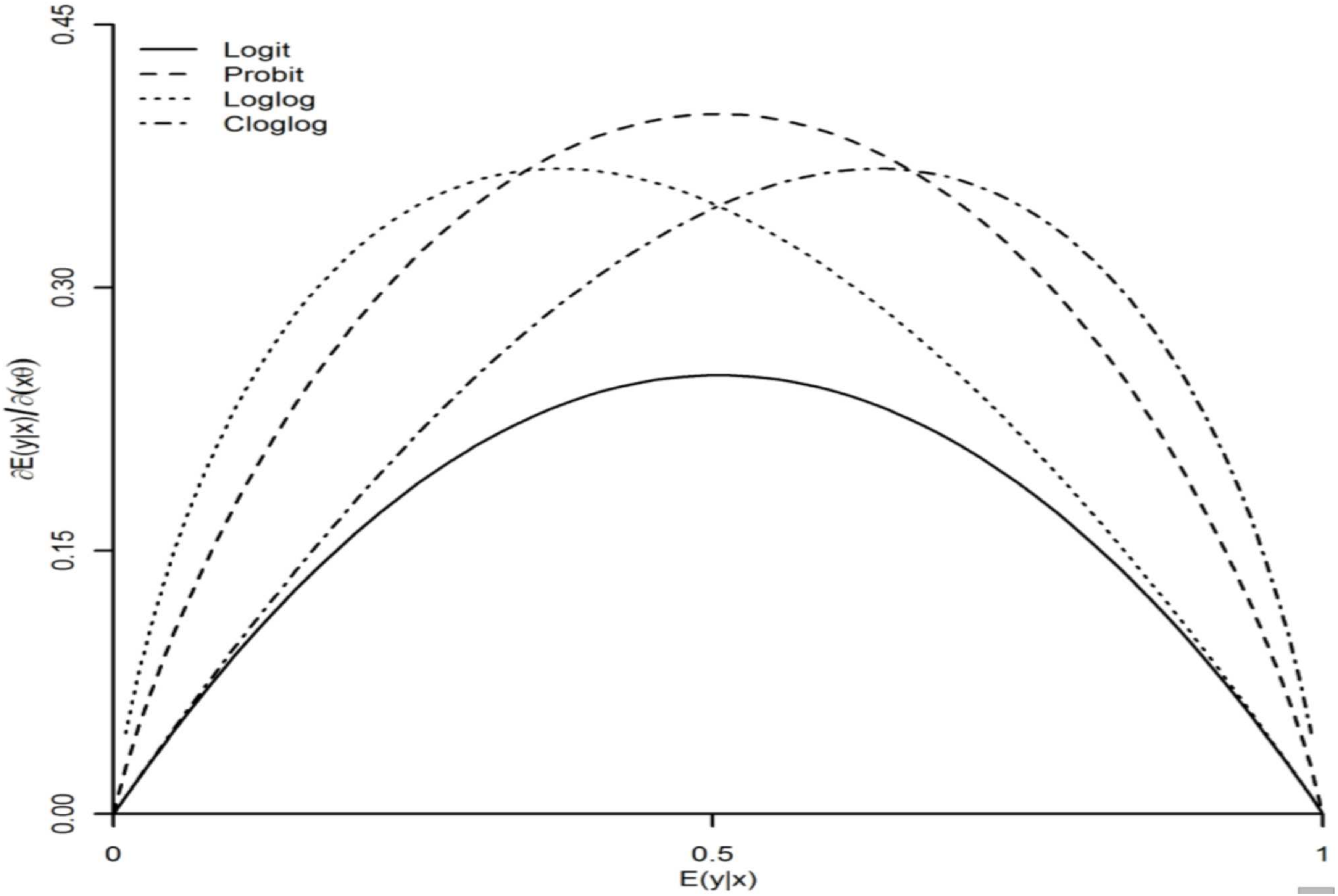
- According to the specification of G, different the resultant model – examples:

  - 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: $G(x_i'\beta) = 1 - e^{-e^{x_i'\beta}}$


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

Partial effects:

- $\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j g(x_i'\beta)$, 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: $g(x_i'\beta) = \phi(x_i'\beta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(x_i'\beta\right)^2}{2}}$
  - 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)

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 $Y \in]0,1[$

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

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
  - $H_1(Y_i)$ 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:

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

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 $Y_{im} \in]0,1[$

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); see Hausman & Leonard (1997)
- Random effects (assumes $\alpha_i \sim N(0,\sigma_\alpha^2)$); see Papke & Wooldridge (2008)
- Fixed effects (based on linear or exponential transformations); see Ramalho & Ramalho (2017)

Stata:
estimator based on quasi mean difference
xtpoisson H(Y) X_1 ... X_k, fe

Control function approach

- Implement the two steps (use a boostrap variance in the second step)

Exponential-fractional conditional mean models (fractional dependente variables)

- Moment condition (Ramalho & Ramalho, 2016)

$$E\left[\frac{H_1(Y)}{exp(x'\beta)} - 1|Z\right] = 0$$

where

- Logit: $H_1(Y_i) = \frac{Y_i}{1-Y_i}$
- Cloglog: $H_1(Y_i) = -ln(1 - Y_i)$

Stata
ivpoisson gmm H1(Y) $(X_1 = IV_A \ldots IV_M)$ X_2 … X_k, multiplicative