Modelos não lineares

Tópicos:

- Estimação e inferência em modelos não lineares
- Modelos para dados binários
- Modelos para excesso de zeros

Motivação para o uso de modelos não lineares:

- Frequentemente a variável dependente é discreta e/ou limitada, não sendo o modelo linear o mais apropriado

Modelos não lineares

Quantidades de interesse:

- Modelos lineares:
  - $E(Y|X)$

- Modelos não lineares:
  - $E(Y|X)$
  - $Pr(Y|X)$
  - Em alguns casos, também outras quantidades:
    - Exemplo: quando $Y \geq 0$, com muitos zeros, pode ser interessante estimar:
      - » $Pr(Y = 0|X)$
      - » $E(Y|X, Y > 0)$

Modelos não lineares

Efeitos parciais:

- Modelos lineares:
  - Modelo: $E(Y|X) = X\beta$
  - Efeitos: $\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \beta_j$

- Modelos não lineares:
  - Modelo:
    - $E(Y|X) = G(X\beta)$
    - $Pr(Y|X) = F(X\beta)$
  - Efeitos: $\Delta X_j = 1 \Longrightarrow$
    - $\Delta E(Y|X) = \frac{\partial E(Y|X)}{\partial X_j} = \frac{\partial G(X\beta)}{\partial X_j} = \beta_j \frac{\partial G(X\beta)}{\partial X\beta} = \beta_j g(x_i'\beta)$
    - $\Delta Pr(Y|X) = \frac{\partial Pr(Y|X)}{\partial X_j} = \frac{\partial F(X\beta)}{\partial X_j} = \beta_j \frac{\partial F(X\beta)}{\partial X\beta} = \beta_j f(x_i'\beta)$

Modelos não lineares

- Os efeitos parciais podem-se comparar para diferentes modelos mas os β não
- Contudo, como $\frac{\partial G(X\beta)}{\partial X\beta} > 0$ e $\frac{\partial F(X\beta)}{\partial X\beta} > 0$:
  - β_j informa sobre o sinal do efeito parcial
  - A significância individual do efeito pode ser testada com $H_0 : \beta_j = 0$
- Para calcular a magnitude dos efeitos parciais:
  - Calcular o efeito para cada individuo e fazer a média no final: efeito parcial médio
  - Substituir x pela sua média: efeito parcial avaliado na média
  - *Substituir* x por valores especificos

Stata
(after estimating the model)
margins, dydx($varlist$)
margins, dydx($varlist$) atmeans
margins, dydx($varlist$) at(...)

Modelos não lineares

Estimação por máxima verosimilhança:

Especifica-se:

- $E(Y|X) = G(X\beta)$
- $Pr(Y|X) = F(X\beta)$

Problema de optimização

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

Propriedades assimptóticas dos estimadores MV:

- Consistencia
- Eficiencia
- Normalidade

Modelos não lineares

Cálculo das variância dos parametros:

- Standard
- Cluster-robust $\longrightarrow$ dados de painel
- Bootstrap

Testes classicos:

- Rácio de verosimilhanças, "Likelihood Ratio" (LR)
- Wald
- Score/LM

Modelos não lineares

Teste para a significância conjunta:

- Modelos:
  - Restrito: $L_R(\beta_0 + \beta_1 x_1 + \cdots + \beta_g x_g)$
  - Não restrito $L_F(\beta_0 + \beta_1 x_1 + \cdots + \beta_g x_g + \beta_{g+1} x_{g+1} + \cdots + \beta_k x_k)$

- Hipoteses:

  H_0: $\beta_{g+1} = \cdots = \beta_k = 0$ (escolha do modelo restrito)

  H_1: $\text{No } H_0$ (escolha do modelo não restrito)

Modelos não lineares

- Teste LR test:

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

  - Disponível na maioria dos packages
  - Há que estimar dois modelos

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

- Teste de Wald :

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

onde $\hat{\beta}_D = \left(\hat{\beta}_{g+1}, \ldots, \hat{\beta}_k\right)$ se baseia em $LL_F(X\beta_F)$

  - Se H_0: $\beta_j = 0$, W simplifica-se:

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

  - Disponível na maioria dos packages
  - Apenas o modelo não restrito é estimado

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

Modelos não lineares

Testes de especificação

1. Teste RESET:

- Estimar o modelo em teste:

$$Pr(Y|X) = F(\beta_0 + \beta_1 x_1 + \cdots \beta_k x_k)$$

- Gerar $(X\hat{\beta})^2, (X\hat{\beta})^3, (X\hat{\beta})^4, \ldots$
- Estimar o modelo auxiliar

$$Pr(Y|X) = F\left[\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma_1 (X\hat{\beta})^2 + \gamma_2 (X\hat{\beta})^3 + \gamma_3 (X\hat{\beta})^4 + \cdots\right]$$

- Aplicar um teste LR / Wald para:

  $H_0: \gamma_1 = \gamma_2 = \gamma_3 = \cdots = 0$ (forma funcional correcta)

  H_1: Não H_0 (forma funcional incorrecta)

Modelos não lineares

2. Teste de Chow:

- Considerando a dummy $D = \begin{cases} 1 \text{ se o individuo pertence a } G_A \\ 0 \text{ se o individuo pertence a } G_B \end{cases}$
- Estimar o modelo 'duplicado':

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

- Aplicar um teste LR / Wald para:

  $H_0{:}\, \gamma_0 = \gamma_1 = \cdots = \gamma_k = 0$ (não existe quebra estrutural)

  H_1: Não H_0

Modelos binários

- $Y \in \{0,1\}$
- Exemplos:
  - Y=1 se obteve crédito bancário
  - Y=1 se comprou o produto A
  - Y=1 se incorreu em incumprimento
- $E(Y_i|x_i) = Pr(Y_i = 1|x_i)$:

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

- Função de densidade nos estimadores MV:

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

onde $G(x_i'\beta) = E(Y_i|x_i) = Pr(Y_i = 1|x_i)$ e $0 < G(\cdot) < 1$, já que $Pr(Y_i = 1|x_i)$ está limitada entre 0 e 1

Modelos binários

- Estimação por MV baseada em :

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

- Modelos mais conhecidos:

  - Probit: $G(x_i'\beta) = \Phi(x_i'\beta) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i'\beta)^2}{2}} dx\beta$
  - Logit: $G(x_i'\beta) = \Lambda(x_i'\beta) = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$
  - Cloglog: $G(x_i'\beta) = 1 - e^{-e^{x_i'\beta}}$

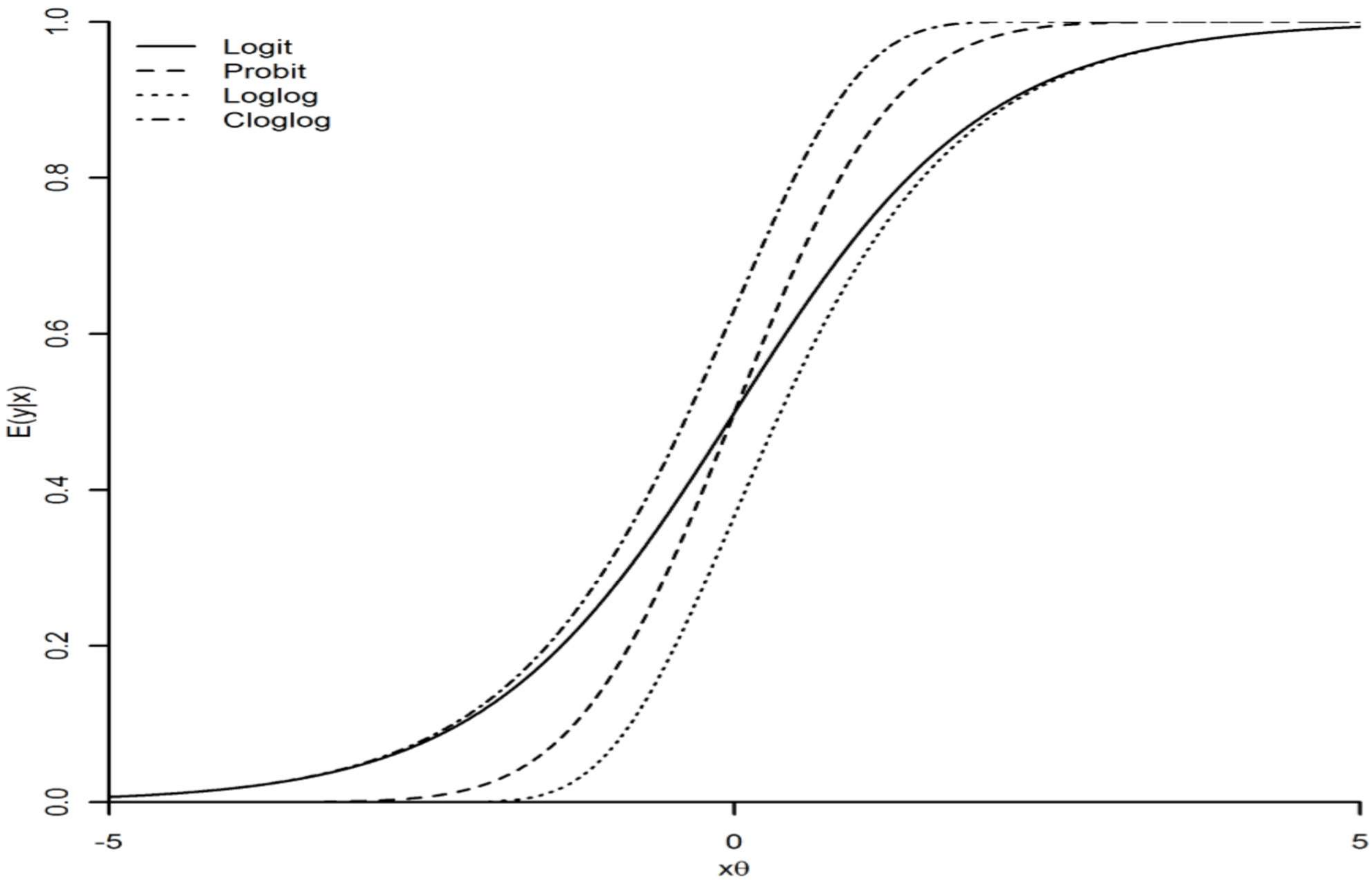
Stata

logit $Y X_1 \ldots X_k$

probit $Y X_1 \ldots X_k$

cloglog $Y X_1 \ldots X_k$

Modelos binários



Modelos binários

Efeitos parciais:

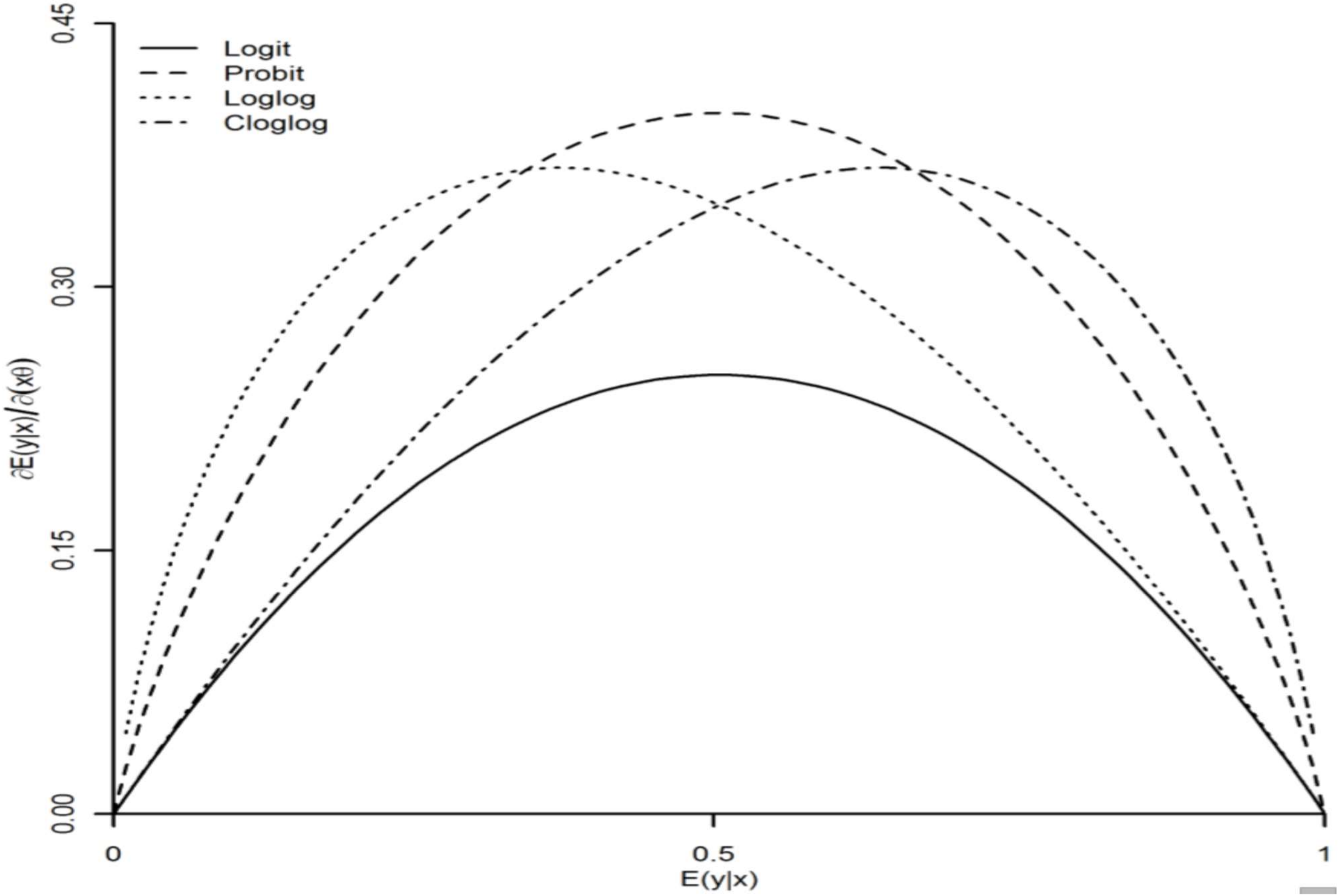
- $\Delta X_j = 1 \Longrightarrow \Delta E(Y|X) = \Delta Pr(Y = 1|X) = \frac{\partial G(X\beta)}{\partial X_j} = \beta_j \frac{\partial G(X\beta)}{\partial (X\beta)} = \beta_j g(x_i'\beta)$, com $g(x_i'\beta)$ dado por:

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

Modelos binários



Modelos binários

Critérios de selecção:

Para "desempatar" modelos não rejeitados pelo RESET , é comum obter a percentagem de previsões correctas:

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

- $\hat{Y}_i = \begin{cases} 1 \text{ if } \widehat{Pr(Y_i = 1 \vert x_i)} \geq 0.5 \\ 0 \text{ if } \widehat{Pr(Y_i = 1 \vert x_i)} < 0.5 \end{cases}$
- % previsões correctas: $(n_{11} + n_{00})/n$
- % 1's correctamente previstos: n_{11}/n_1
- % 0's correctamente previstos: n_{00}/n_0

Stata
(depois de estimar o modelo)
estat classification

Modelos binários

- **Exemplo: Loan application**
- Variável dependente: *aprove* (=1 se obteve crédito à habitação)
- Variáveis explicativas: *hrat* (peso da prestação no rendimento mensal em %), *married* (=1 se casado), *dep* (número de dependentes do agregado familiar)

Exemplo: Loan application

```
. logit approve married dep hrat
Logistic regression                               Number of obs     =      1,986
                                                  LR chi2(3)        =      20.14
                                                  Prob > chi2       =     0.0002
Log likelihood = -729.88535                       Pseudo R2         =     0.0136
------------------------------------------------------------------------------
     approve |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .4848471   .1534378     3.16   0.002     .1841145    .7855796
         dep |  -.1410653   .0642852    -2.19   0.028     -.267062   -.0150686
        hrat |  -.0267698   .0094215    -2.84   0.004    -.0452355    -.008304
       _cons |   2.453316   .2693916     9.11   0.000     1.925318    2.981314
------------------------------------------------------------------------------
. probit approve married dep hrat
Probit regression                                 Number of obs     =      1,986
                                                  LR chi2(3)        =      19.52
                                                  Prob > chi2       =     0.0002
Log likelihood = -730.19351                       Pseudo R2         =     0.0132
------------------------------------------------------------------------------
     approve |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |    .255545   .0812719     3.14   0.002     .0962551     .414835
         dep |  -.0747856   .0345529    -2.16   0.030    -.1425081   -.0070632
        hrat |  -.0134307    .004899    -2.74   0.006    -.0230325   -.0038289
       _cons |   1.397252   .1395427    10.01   0.000     1.123753    1.670751
------------------------------------------------------------------------------
```

Modelos binários

Para o caso do logit:

$$Pr(\widehat{approve} = 1|X\ ) = \Lambda(2.453 + 0.485 married - 0.141 dep - 0.027\ hrat)$$

Pode-se apenas afirmar que a probabilidade de ter um credito aprovado aumenta no caso do individuo ser casado e diminui quando o número de dependentes e a taxa de esforço para pagamento aumentam. Para se ter a magnitude do impacto sobre a probabilidade em estudo, será necessário calcular os efeitos parciais

Para o caso do probit:

$$Pr(\widehat{approve} = 1|X\ ) = \Phi(1.397 + 0.256 married - 0.075 dep - 0.013\ hrat)$$

O comentário anterior aplica-se, apesar da diferença natural da magnitude dos coeficientes

Modelos binários

Restringindo o interesse ao logit, por exemplo:

Efeitos parciais

```
.  margins, dydx ( married dep hrat )
Average marginal effects                              Number of obs     =      1,986
Model VCE     : OIM
Expression    : Pr(approve), predict()
dy/dx w.r.t. : married dep hrat
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx    Std. Err.       z     P>|z|      [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |    .0516943   .0163988     3.15    0.002      .0195532    .0838354
         dep |   -.0150404   .0068616    -2.19    0.028     -.0284888   -.0015919
        hrat |   -.0028542   .0010057    -2.84    0.005     -.0048254    -.000883
------------------------------------------------------------------------------
```

Modelos binários

Teste LR para a significância conjunta

```
. logit approve
Logistic regression                               Number of obs     =      1,986
                                                  LR chi2(0)        =      -0.00
                                                  Prob > chi2       =          .
Log likelihood = -739.95364                       Pseudo R2         =    -0.0000
------------------------------------------------------------------------------
     approve |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   1.965621   .0683551    28.76   0.000     1.831647    2.099594
------------------------------------------------------------------------------
```

- $LR = 2(-729.885 + 739.954) = 20.14$, a 5% $\chi^2_3 = 7.814$

Modelos binários

Teste LR para a significância conjunta de *dep* e *hrat*

```
. logit approve married
Logistic regression                             Number of obs     =      1,986
                                                LR chi2(1)        =       7.08
                                                Prob > chi2       =     0.0078
Log likelihood = -736.41424                     Pseudo R2         =     0.0048
------------------------------------------------------------------------------
     approve |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .3743725   .1394271     2.69   0.007     .1011004    .6476446
       _cons |   1.731135   .1074245    16.11   0.000     1.520587    1.941683
------------------------------------------------------------------------------
```

- $LR = 2(-729.885 + 736.414) = 13.30$, a 5% $\chi_2^2 = 5.99$

Modelos binários

Teste RESET

```
. quietly logit approve married dep hrat
. predict xb ,xb
. gen xb2=xb^2
. gen xb3=xb^3
. quietly logit approve married dep hrat xb2 xb3

. test xb2 xb3
 ( 1)   [approve]xb2 = 0
 ( 2)   [approve]xb3 = 0

           chi2(  2) =   11.57
         Prob > chi2 =    0.0031
```

Obtenção de probabilidade de sucesso para cada individuo

```
. quietly logit approve married dep hrat

. predict P1, pr
```

Modelos para excesso de zeros

Objectivo: descrever casos onde a variável dependente assume

- O mesmo valor para muitos indivíduos: $Y_i = 0$
- Um valor positivo para os restante indivíduos

Exemplos:

- Gastos em bens duradouros, álcool, …
- Horas de trabalho
- % de dívida de longo prazo em PME's

Modelos estudados:

- Tobit: um modelo único descreve os 0's e os valores positivos
- Modelos a 2 partes: envolvem dois modelos independentes para explicar os 0's e os valores positivos

Especificação:

- Modelo latente: $Y_i^* = x_i'\beta + u_i$
  - Em lugar de Y_i^* observa-se:

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

com $u_i \sim N(0, \sigma^2)$

– $Pr(Y_i = 0|x_i) = Pr(Y_i^* \leq 0|x_i) = Pr(x_i'\beta + u_i \leq 0|x_i) = Pr(u_i \leq -x_i'\beta|x_i) = Pr\left(\frac{u_i}{\sigma} \leq -\frac{x_i'\beta}{\sigma}\middle| x_i\right) = \Phi\left(-\frac{x_i'\beta}{\sigma}\right) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)$

- Descrição dos dados: $f(y_i|x_i) = \begin{cases} 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) & \text{if Y} = 0 \\ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y_i - x_i'\beta\right)^2}{2\sigma^2}} & \text{if } Y > 0 \end{cases}$

Estimação:

- Metodo: MV

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

- Parametros a estimar: β and σ

- Função de log-verosimilhança:

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

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

Quantidades de interesse:

- Média condicional dado que Y_i é positivo:

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

onde $\lambda\left(\frac{x_i'\beta}{\sigma}\right) = \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}$ é designado de rácio de Mills

- Probabilidade de observar valores positivos de Y_i:

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

- Média condicional global:

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

Efeitos parciais:

- $\Delta X_j = 1 \Longrightarrow$
  - $\Delta E(Y_i|x_i, Y_i > 0) = \beta_j \left\{1 - \lambda\left(\frac{x_i'\beta}{\sigma}\right)\left[\frac{x'_i\beta}{\sigma} + \lambda\left(\frac{x_i'\beta}{\sigma}\right)\right]\right\}$
  - $\Delta Pr(Y_i > 0|x_i) = \frac{\beta_j}{\sigma}\phi\left(\frac{x_i'\beta}{\sigma}\right)$
  - $\Delta E(Y_i|x_i\ ) = \beta_j\Phi\left(\frac{x_i'\beta}{\sigma}\right)$
- O sinal é igual para todos. Naturalmente, a magnitude difere

Modelos para excesso de zeros
2 partes (tópico extra)

Especificação:

- Parte 1– modelo binário:

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

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

- Parte 2– exponential, linear, ...

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

- Média condicional:

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

Modelos para excesso de zeros
2 partes (tópico extra)

Estimação:

Cada parte é estimada separadamente:

- Na parte 1: usa-se a totalidade da amostra
- Na parte 2: usa-se a subamostra com $Y_i > 0$
- Podem ser usadas as mesmas ou diferentes variáveis explicativas em cada parte

Efeitos parciais:

- $\Delta Pr(d_i = 1|x_i) = \Delta Pr(Y_i > 0|x_i)$
- $\Delta E(Y_i|x_i, d_i = 1) = \Delta E(Y_i|x_i, Y_i > 0\ )$
- $\Delta E(Y_i|x_i\ ) = \Delta Pr(d_i = 1|x_i) E(Y_i|x_i, d_i = 1) + Pr(d_i = 1|x_i) \Delta E(Y_i|x_i, d_i = 1)$