MODELOS DE RESPOSTA BINÁRIA

Existe uma variedade de comportamentos econômicos onde as aproximações contínuas não são boas, por exemplo:

(a) Indivíduos: escolha de emprego; número de filhos; se compra uma casa; se é inadimplente; se participa da escola; se opta casar.
(b) Empresas: se construir uma planta, e se sim, qual localização; qual produto produzir; se para a dívida (se dá o default); adquirem-se outras empresas.

Para isso precisamos de modelos de probabilidade que aproximem o verdadeiro processo gerador dos dados. Nós nos concentraremos no caso em que a variável explicada, y, é uma resposta binária.
Por exemplo, y=1 se a pessoa é inadimplente, y=0 caso contrário.

Em econometria tratamos estas variáveis como dummy.

A variável dummy é um simples exemplo de variável aleatória (ξ), o qual é chamada de função indicadora de um conjunto A:

$$\xi = I_A(x)$$

onde,

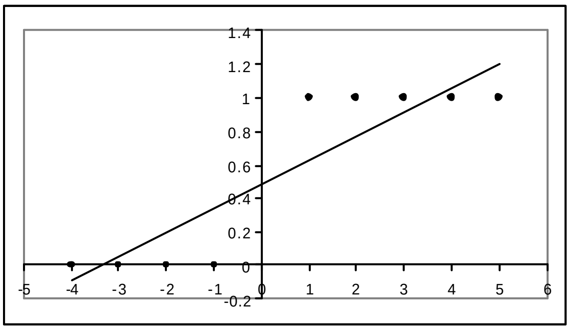
$$I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

São utilizadas para diversos fins:
a) Reconhecer diferenças no intercepto
b) Reconhecer diferenças na inclinação
c) Sazonalidade
d) Variáveis qualitativas: diferenças em idade, sexo, escolaridade,…
e) Comprovam o teste de estabilidade
f) Como variável dependente: Logit e Probit

<u>Variável dummy dependente</u>

Em geral trabalhamos com modelos em que as variáveis explicativas (x) são dummies, agora veremos modelos em que variáveis explicadas (y) são dummies.

A idéia de fazer uma estimativa por OLS, tentando aproximar uma linha de regressão linear, temos que os valores previstos podem ficar fora do intervalo [0,1] e assim os erros de previsão serão grandes.



De forma a evitar estes problemas faremos as estimativas via modelos Probit e Lobit, o qual chamarei de modelos de resposta binária.

Assuma: $y_i^* = a + \sum_{j=1}^{k} b_j x_{ij} + e_i$

onde, y* é não observada (variável latente), isto é, poderia ser definida como a propensão de uma empresa entrar em default.

O que nós observamos é uma variável dummy y_i definida por:

$$y_i = \begin{cases} 1 & \text{se } y^* > 0 \\ 0 & \text{caso contrário} \end{cases}$$

Se multiplicarmos y* por qualquer constante positiva não mudamos y_i com isso podemos assumir $V(\varepsilon_t)=1$. Isto fixa a escala de y*, então:

$$P_i = \Pr(y_i = 1) = \Pr[a + \sum_{j=1}^{k} b_j x_{ij} + e_i > 0]$$

$$= \Pr[e_i > -(a + \sum_{j=1}^{k} b_j x_{ij})]$$

$$= 1 - F[-(a + \sum_{j=1}^{k} b_j x_{ij})]$$

onde, F é a função distribuição cumulativa do erro (ε), contínua, estritamente crescente que toma valores reais e retorna a valores entre zero e um.

Se a distribuição de ε é simétrica, desde que 1-F(-Z)=F(Z) podemos escrever:

$$P_i = F[a + \sum_{j=1}^{k} b_j x_{tj}] \qquad (1)$$

A escolha da função F determina o tipo de modelo binário (Probit e Logit).

Para o modelo de probabilidade linear (OLS):

F(z)=z onde $z = a + \sum_{j=1}^{k} b_j x_{tj}$

é a função identidade, no qual significa que a resposta probabilística não necessariamente estará entre 0 e 1 para todo x e ß.

Nesta seção assumimos que F(·) toma valores no intervalo aberto unitário: 0<F(z)<1 para todo z∈R.

O modelo **Probit** é um caso especial da equação (1) com:

$$F(z) = \Phi(z) = \int_{-\infty}^{z} f(u)du$$

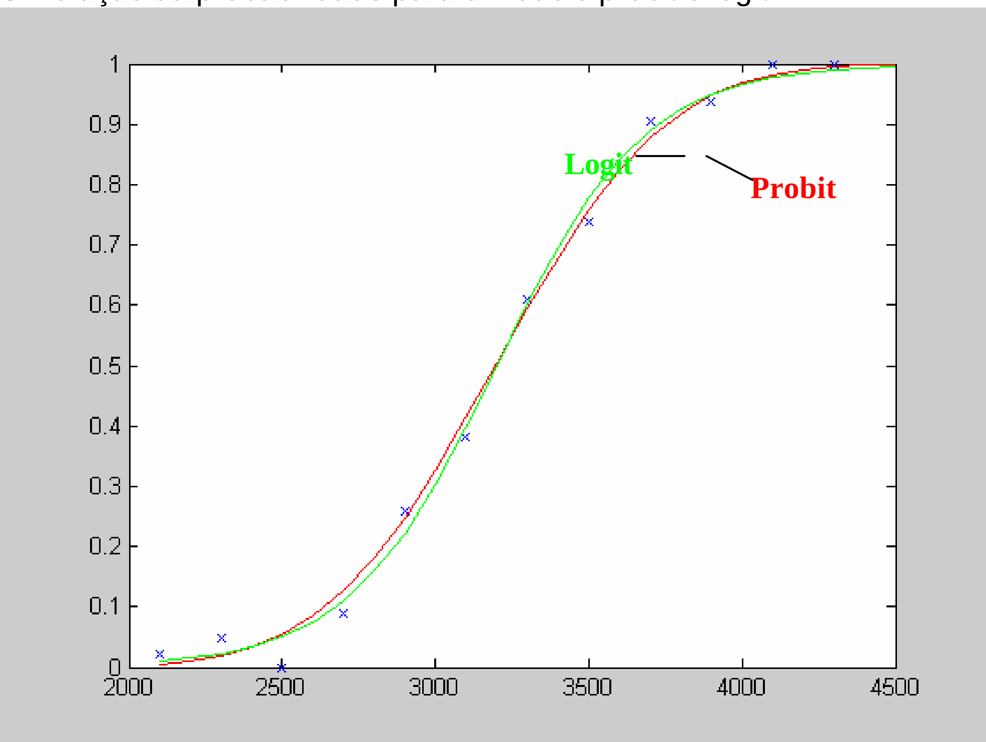
onde ϕ(z) é a densidade da normal padrão

$$f(z) = \frac{1}{\sqrt{2p}} \exp\left(-\frac{z^2}{2}\right)$$

O modelo **Logit** é um caso especial da equação (1) com:

$$F(z) = \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)}$$

Simulação da probabilidade para o modelo probit e logit



Pela diferença dos dois métodos, temos os resultados podem ser diferentes exceto para amostras muito grandes. Entretanto, as estimativas dos parâmetros α e os β_i dos dois métodos não são comparáveis.

Dada tal especificação, podemos estimar os parâmetros deste modelo usando o **método da máxima verossimilhança**. Desde que os dados observados y_i são realizações de um processo binomial com probabilidade P_i a função de verossimilhança é dada por:

$$L = \prod_{y_t=1} P_i \prod_{y_t=0} (1 - P_i)$$

Transformando para a forma de log, temos:

$$\log L(a, b) = \sum_{i=0}^{n} y_i \log(F[a + \sum_{j=1}^{k} b_j x_{tj}]) + (1 - y_i)\log(1 - F[a + \sum_{j=1}^{k} b_j x_{tj}])$$

As condições de primeira ordem para esta função log da verossimilhança são não lineares, tal que para estimar os parâmetros é necessário um procedimento iterativo (otimização numérica).

Interpretação dos modelos de resposta binária

A interpretação dos parâmetros não é tão simples como o básico modelo de regressão OLS (onde podemos achar ß como uma derivada parcial). Isto é por causa da não linearidade.

Para aplicar os modelos probit e logit, é importante saber como interpretar os $ß_j$ no caso de variáveis explicativas contínua e discreta.
Primeiro, se x_j é contínua

$$\frac{\partial p(x)}{\partial x_j} = f(xb) b_j$$

onde

$$f(z) = \frac{dF}{dz}(z)$$

Se F(·) é uma função distribuição cumulativa (cdf) estritamente crescente, como nos casos do probit e logit, f(z)>0 para todo z. Então, o sinal do efeito é dado pelo sinal do $ß_j$.

No caso probit
f(z)=ϕ(z)

No caso logit

$$f(z) = \frac{\exp(z)}{(1 + \exp(z))^2}$$

Fazer o caso z =0

Se x_k é uma variável explicativa binária, então o efeito parcial da troca x_k de zero para um, mantendo todas as outras fixas, é simplemente:

$$F(b_1 + b_2 x_2 + \ldots + b_{k-1} x_{k-1} + b_k) - F(b_1 + b_2 x_2 + \ldots + b_{k-1} x_{k-1})$$

Ver o caso para outros tipos de variáveis discretas

É trivial incluir formas funcionais padrão nas variáveis explicativas. Por exemplo, no modelo:

$$P(y = 1 | \mathrm{z}) = F(b_0 + b_1 z_1 + b_2 z_1^2 + b_3 \log(z_2) + b_4 z_3)$$

Ver o efeito parcial de z_1 e z_2

Para fazer comparável as estimativas dos modelos Logit e Probit, podemos usar o fato que aproximações da trocas na probabilidade são:

$$\Delta\hat{P}(y = 1 | \mathrm{z}) \approx [f(\hat{b}_0 + \hat{b}z)\hat{b}_j]\Delta z_j$$

para pequenas trocas em z_j, então, o fator f pode ser usado para ajustar cada beta.

Assim, podemos comparar as magnitudes das inclinações estimadas do probit e logit. Tomando para o probit f(0)≈0.4 e para o logit f(0)≈0.25. Então, para fazer comparável as estimativas da inclinação para logit e probit, podemos multiplicar as estimativas do probit por 0.4/0.25=1.6 ou a do logit por 0.625. No trabalho de Amemiya[1], sugere multiplicar as estimativas Logit por 0.625 e também sugere a comparação entre OLS e Probit.

<u>Regra de bolso</u> para tornar os modelos comparáveis:

1) α e β (Logit) × 0.625 = α e β (Probit)
2) α e β (OLS) × 2.5 e α (OLS) – 1.25 = α e β (Probit)

<u>Exemplo 10</u>: Medir o efeito de execuções do governo, com o seguinte modelo:

$$D_1 = a + b_1 T + b_2 Y + b_3 LF + b_4 NW$$

onde, D_1 = 1 se o governo faz execuções e zero caso contrário
T = mediana de tempo em meses de o assassino ser condenado
Y = mediana da renda familiar
LF = taxa de participação na força de trabalho
NW = proporção da população não branca
Para comparar faremos três estimativas, por OLS, Logit e Probit.

```
>> ols1 = ols(ex10(:,1),[ones(44,1),ex10(:,2:5)]);
>> prt_reg(ols1)
```

[1] T. Amemiya, "Qualitative Response Model: A Survey", Journal of Economic Literature, 1981, p. 1488.

```
Ordinary Least-squares Estimates
R-squared      =    0.2802
Rbar-squared   =    0.2063
sigma^2        =    0.1321
Durbin-Watson  =    2.0779
Nobs, Nvars    =     44,     5
***************************************************************
Variable       Coefficient      t-statistic    t-probability
variable 1        2.733391         2.105437         0.041744
variable 2        0.000834         0.866134         0.391714
variable 3        0.535329         2.254490         0.029857
variable 4       -0.061771        -2.145844         0.038167
variable 5        2.583121         3.664107         0.000737

>> log1 = logit(ex10(:,1),[ones(44,1),ex10(:,2:5)]);
>> prt_reg(log1)

Logit Maximum Likelihood Estimates
McFadden R-squared     =    0.5914
Estrella R-squared     =    0.5963
LR-ratio, 2*(Lu-Lr)    =   26.3691
LR p-value             =    0.0000
Log-Likelihood         =   -9.1076
# of iterations        =     11
Convergence criterion  =    6.456733e-011
Nobs, Nvars            =     44,     5
# of 0's, # of 1's     =      9,    35
***************************************************************
Variable       Coefficient      t-statistic    t-probability
variable 1       16.566989         0.843565         0.404057
variable 2        0.016516         1.718467         0.093642
variable 3        9.131548         1.807101         0.078466
variable 4       -0.715389        -1.492673         0.143569
variable 5       85.361600         2.380780         0.022257

>> prob1 = probit(ex10(:,1),[ones(44,1),ex10(:,2:5)]);
>> prt_reg(prob1)

Probit Maximum Likelihood Estimates
McFadden R-squared     =    0.5955
Estrella R-squared     =    0.6004
LR-ratio, 2*(Lu-Lr)    =   26.5509
LR p-value             =    0.0000
Log-Likelihood         =   -9.0167
# of iterations        =     10
Convergence criterion  =    1.6172748e-007
Nobs, Nvars            =     44,     5
# of 0's, # of 1's     =      9,    35
***************************************************************
Variable       Coefficient      t-statistic    t-probability
variable 1       10.267388         0.978093         0.334057
variable 2        0.009421         1.858119         0.070713
variable 3        5.551099         1.973971         0.055499
variable 4       -0.436545        -1.702430         0.096634
variable 5       50.248741         2.501305         0.016681
```

Para tornar os modelos comparáveis utilizaremos a regra de bolso de Amemiya.

```
>> ols_probit=[(ols1.beta(1,1)*2.5)-1.25;ols1.beta(2:5,1)*2.5];
>> logit_probit=log1.beta*0.625;
>> coef=[ols_probit, logit_probit, prob1.beta]

coef =

    5.5835   10.3544   10.2674
    0.0021    0.0103    0.0094
    1.3383    5.7072    5.5511
   -0.1544   -0.4471   -0.4365
    6.4578   53.3510   50.2487
```

Como podemos perceber o modelo por OLS ficou com os coeficientes bastante diferentes dos modelos Logit e Probit, enquanto que estes dois últimos ficaram muito semelhantes.

Testes de especificação

a) Autocorrelação:
Pelo plote da função de autocorrelação (autocorr) dos três modelos não há evidência dos erros serem correlacionados, se os dados são em cross-section não precisamos fazer este teste.

b) Multicolinearidade:
Calculando os coeficientes de correlação simples das variáveis (corrcoef) também não evidenciamos problema de multicolinearidade, mas cabe ressaltar, que é importante fazer o teste com as correlações múltiplas R^2_i.

c) Heterocedasticidade:
Para testar heterocedasticidade quando a variável explicada é binária, como os resíduos tomam valores $\hat{P}$ ou $1-\hat{P}$ para cada observação, não é suficiente apenas fazer um teste de White, em que testávamos a hipótese H_0: $V(e_t)=s^2$ (homocedasticidade) contra H_1: $V(e_t)=s^2 z_t^2$ (heterocedasticidade)

No modelo binário (Probit) temos:

$$y_i^* = a + \sum_{j=1}^{k} b_j x_{ij} + e_i \quad \text{sendo} \quad e_i \text{ » } IN(0,1)$$

Mas como observamos apenas o y sendo 0 ou 1, podemos escrever o modelo como:

Detonate, $Z_i = a + \sum_{j=1}^{k} b_j x_{tj}$

Vimos que: $P_i = \Pr(y_i = 1 | Z_i) = 1 - F[-Z_i] = F(Z_i)$

$E(y_i = 1 | Z_i) = 1 * \Pr(y_i = 1 | Z_i) + 0 * \Pr(y_i = 0 | Z_i) = \Pr(y_i = 1 | Z_i)$

$y_i = F(Z_i) + u_i$ sendo F a distribuição cumulativa de ε e υ é um resíduo representando o desvio da binária y de sua média condicional. Então:

$E(u_i | Z_i) = E(y_i - F(Z_i)) = 0$

$V(u_i | Z_i) = E(u_i)^2 = F(Z_i)(1 - F(Z_i))$

Uma especificação mais geral é que os erros sejam heterocedástico:

$y_i^* = a + \sum_{j=1}^{k} b_j x_{ij} + e_i$ sendo $e_i \sim N(0, \exp(2 z_i' g))$

onde z_i é uma das variáveis explicativas e γ um parâmetro não conhecido.

Se γ=0 temos que os erros são homocedásticos, assim o teste conforme Davidson e Mackinnon (1993) é feito por testar as seguintes hipóteses:

H_0: γ=0 (homocedástico)

H_1: γ≠0 (heterocedástico)

Para testar as hipóteses, montamos a seguinte regressão dos resíduos:

Como vimos que υ_i tem variância $F(Z_i)(1 - F(Z_i)$ dividimos ambos os lados da regressão por $(V(u_i | Z_i))^{-1/2} = (F(Z_i)(1 - F(Z_i))^{-1/2} = (\hat{y}_i (1 - \hat{y}_i))^{-1/2}$.

$$\frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i (1 - \hat{y}_i)}} = a \frac{f(-Z_i)}{\sqrt{\hat{y}_i (1 - \hat{y}_i)}} + \sum_{j=1}^{k} b_j \frac{f(-Z_i)}{\sqrt{\hat{y}_i (1 - \hat{y}_i)}} x_{ij} + c \frac{f(-Z_i)(-Z_i)}{\sqrt{\hat{y}_i (1 - \hat{y}_i)}} z_i$$

onde $\hat{y}$ é o y estimado para cada observação, $f(-Z_i)$ é a densidade da distribuição (pdf).

A estatística do teste é dada pela soma do quadrado dos valores estimados (resíduos padronizados). Este teste tem distribuição χ^2 com graus de liberdade igual ao número de variáveis z.

Uma outra boa referência é o Greene[2]

Exemplo 11: De acordo com Greene. Modelo para analisar a efetividade de um novo método de ensinar economia. O modelo é:

$GRADE = a + b_1 GPA + b_2 TUCE + b_3 PSI$

onde, GRADE = 1 se o estudante melhorou seu coeficiente 0 caso contrário.

GPA = coeficiente médio

TUCE = conhecimento do material

PSI = 1 se o estudante foi exposto ao novo método e 0 caso contrário.

```
>> prob1 = probit(ex11(:,1),[ones(32,1),ex11(:,2:4)]);
>> p_hat=prob1.yhat;
>> xb=-[ones(32,1),ex11(:,2:4)]*prob1.beta;
>> brmr_y=prob1.resid./sqrt(p_hat.*(1-p_hat));
```

[2] Greene, W. Econometric Analysis. 3º ed. Prentice-Hall, 1997.

```
>> fac=normpdf(-xb)./sqrt(p_hat.*(1-p_hat));
>> brmr_x=[fac,ex11(:,2).*fac,ex11(:,3).*fac,ex11(:,4).*fac];
>> oslh=ols(brmr_y,[brmr_x,ex11(:,4).*(-xb).*fac]);
>> brmr_yf=oslh.yhat;
>> lm_test=sum((brmr_yf).^2);
lm_test =

   1.5480
>> p_val = 1-chi2cdf(lm_test,1);
p_val =

   0.2134
```

Acima temos que não rejeito H0 (homocedástico) com probabilidade = 78.66%

Exercício: Testar heterocedasticidade para o exemplo 10.

d) Inclusão de variáveis irrelevantes:
Considere o modelo:
$P(y=1 \mid \mathrm{x}, \mathrm{z}) = G(x\beta + z\gamma)$
onde x é $1 \times$ K e z é $1 \times$ Q. Queremos testar na hipótese nula:
H_0: $\gamma=0$ (Q restrições de exclusões).
O elemento de z pode ser funções de x, tais como quadráticas – neste caso o teste é somente da forma funcional. Ou z pode ser variáveis explicativas adicionais.
Podemos fazer os teste padrão Wald, LR e LM. O teste LR pode ser facilmente computado, estimando o log-likehoood do modelo restrito e não restrito.
O teste LM é atrativo se o modelo não restrito é difícil de ser estimado. A estatística LM usando a hessiana esperada A é numericamente idêntica ao seguinte:
1) Defina $\hat{u}_i = y_i - G(x_i\hat{\beta})$, $\hat{G}_i = G(x_i\hat{\beta})$, $\hat{g}_i = g(x_i\hat{\beta})$ todas são obtidas depois de estimar o modelo sem z.
2) Usar todas N observações para rodar regressão OLS

$$\frac{\hat{u}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} \, on \, \frac{\hat{g}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} x_i, \; \frac{\hat{g}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} z_i$$

A estatística LM é igual a soma do quadrado dos valores estimados desta regressão, com distribuição χ^2 com Q graus de liberdade. Um teste que é assintoticamente (mas não numericamente) equivalente é NR^2 onde R^2 é o R-quadrado não centrado da regressão acima.
O teste LM é atrativo com relação ao LR se z tem uma dimensão grande, pois o modelo probit é difícil de estimar para muitas variáveis explicativas.

e) Testando não linearidade em β:
O teste de Wald é indicado, pois o estimador não restrito de β no qual é o modelo Logit e Probit, é fácil estimar.
H_0: c(β)=0 onde c(β) é um vetor Q × 1 de possíveis funções não lineares. A estatística é:

$$W = c(\hat{b})'[\nabla_b c(\hat{b})\hat{V}\nabla_b c(\hat{b})']^{-1} c(\hat{b})$$

onde $\nabla_b c(\hat{b})$ é o Jacobiano Q × K de c(β).

Teste de performance

As estimativas dos modelos nada nos diz sobre qual modelo utilizar ou qual modelo é melhor, então para avaliar o modelo com melhor ajuste, usaremos medidas tipo R^2, que são as seguintes:

$$\text{McFadden } R^2 = 1 - \frac{\log L_{UR}}{\log L_R}$$

onde L_{UR} é o máxima da função de verossimilhança[3] do modelo e L_R é o máximo da função de verossimilhança do modelo com as restrições dos β_i=0 para i=1,2,…,k, ou seja, é rodar a regressão com a constante apenas.

Uma outra medida é R^2 é dada por: $R^2 = 1 - \left(\frac{L_{UR}}{L_R}\right)^{2/n}$, onde n=nº observações.

Para os três modelos acima o R^2 de McFadden já é calculado no programa Probit e Logit, sendo que temos que calcular para o OLS.
No exemplo 10 temos:

```
>> L_OLS_UR=-44/2*((1+log(2*3.141593))+log(sum(ols1.resid.^2)/44))
L_OLS_UR =  -15.2532

>> ols1_r = ols(ex10(:,1),ones(44,1));
>> L_OLS_R=-44/2*((1+log(2*3.141593))+log(sum(ols1_r.resid.^2)/44))
L_OLS_R = -22.4856
>> R2_MacFadden=1-(L_OLS_UR/L_OLS_R)
R2_MacFadden =  0.3216

>> log1.r2mf
ans = 0.5914

>> prob1.r2mf
ans = 0.5955
```

[3] A função log da máxima verossimilhança é definida por: $\log L = -\frac{n}{2}\left(1 + \log(2p) + \log(\sum e^2 / n)\right)$

Pelo R^2 de MacFadden o melhor modelo é o Probit e o OLS apresenta uma performance pobre.

Além do teste de performance de McFadden R^2 podemos ver o poder de previsão do modelo. Neste caso temos que:

1) Reestimar o modelo com N-n observações
2) Prever as n observações, só que neste caso é preciso considerar se $G(x_i\hat{b}) > 0.5$ prevemos y_i sendo a unidade e se $G(x_i\hat{b}) \leq 0.5$ y_i sendo zero.
3) O percentual de vezes que o valor previsto de y foi o observado é o percentual previsto corretamente.

Bibliografia

MADDALA, G. S. *Introduction to Econometrics*. Prentice Hall, 2º ed, 1992.

DAVIDSON, R. and MACKINNON, J. *Estimation and inference in econometrics*. Oxford, 1993.

LESAGE, James P. Econometrics Toolbox. http://www.spatial-econometrics.com/.

WOOLDRIDGE, J. M. Econometric Analysis of Cross-Section and Panel Data. MIT Press, 2002.