Econometrics - The LPM, probit, logit models

Master 1 Semestre 2 - EPOLPRO (IEDES)

Jean-Baptiste Guiffard (Univ. Paris-1 Panthéon-Sorbonne)

03 mars 2025

Introduction

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners
  - Unlikely: on a random sample of the population

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners
  - Unlikely: on a random sample of the population
  - $\rightarrow$ Who are those working ?

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners
  - Unlikely: on a random sample of the population
  - $\rightarrow$ Who are those working ?
- OLS models that estimate the differenciated effect on wage of junior college education and of university education

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners
  - Unlikely: on a random sample of the population
  - $\rightarrow$ Who are those working ?
- OLS models that estimate the differenciated effect on wage of junior college education and of university education
  - If returns of junior college are found to be higher: is it because of junior college education per se or because it selects the best students ?

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
    - Likely: on a random sample of earners
    - Unlikely: on a random sample of the population
    - $\rightarrow$ Who are those working ?
- OLS models that estimate the differenciated effect on wage of junior college education and of university education
    - If returns of junior college are found to be higher: is it because of junior college education per se or because it selects the best students ?
    - $\rightarrow$ Who attends junior college education ?

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners
  - Unlikely: on a random sample of the population
  - $\rightarrow$ Who are those working ?
- OLS models that estimate the differenciated effect on wage of junior college education and of university education
  - If returns of junior college are found to be higher: is it because of junior college education per se or because it selects the best students ?
  - $\rightarrow$ Who attends junior college education ?

Introduction

So far, we have notably seen

- OLS models that estimate the link between earnings and education
  - Likely: on a random sample of earners
  - Unlikely: on a random sample of the population
  - $\rightarrow$ Who are those working ?
- OLS models that estimate the differenciated effect on wage of junior college education and of university education
  - If returns of junior college are found to be higher: is it because of junior college education per se or because it selects the best students ?
  - $\rightarrow$ Who attends junior college education ?

$\Rightarrow$ How deal with binary outcomes / choices ?

Modelling binary choices

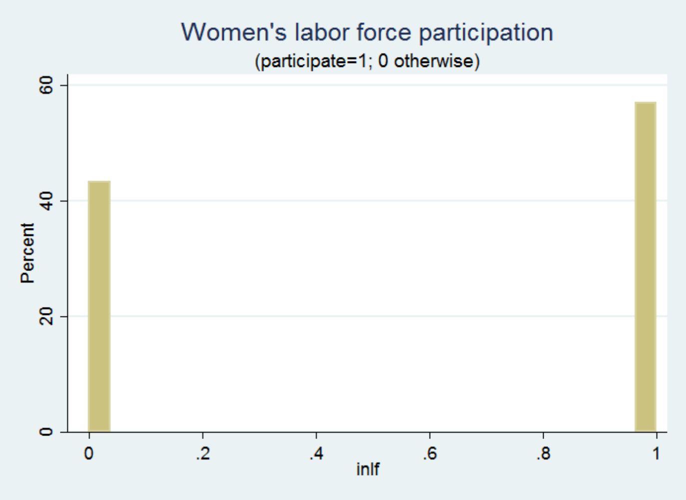
- We can model binary outcomes using OLS models but they do have some limitations

Modelling binary choices

- We can model binary outcomes using OLS models but they do have some limitations
- Most often: we use probit and logit models

The linear probability model (LPM)

Exemple: Women's labor force participation (US, 1975)



The LPM: intuitions (1)

- Descriptively, we find that 56.8% of women are in the labor market

```
. reg inlf

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(0, 752)       =      0.00
       Model |           0         0           .   Prob > F        =         .
    Residual |  184.727756       752  .245648611   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =    0.0000
       Total |  184.727756       752  .245648611   Root MSE        =    .49563

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5683931   .0180617    31.47   0.000     .5329357    .6038505
------------------------------------------------------------------------------
```

The LPM: intuitions (1)

- Descriptively, we find that 56.8% of women are in the labor market
- This is the value of the constant in an OLS model regressing the labor force participation on the constant

```
. reg inlf

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(0, 752)       =      0.00
       Model |           0         0           .   Prob > F        =         .
    Residual |  184.727756       752  .245648611   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =    0.0000
       Total |  184.727756       752  .245648611   Root MSE        =    .49563

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5683931   .0180617    31.47   0.000     .5329357    .6038505
------------------------------------------------------------------------------
```

The LPM: intuitions (2)

Women's labor force participation by presence of children under 6 (US, 1975)

The LPM: intuitions (3)

Women's labor force participation by presence of children under 6 (US, 1975)

inlf	Dkidslt6 0	1	Total
0	231 38.12	94 63.95	325 43.16
1	375 61.88	53 36.05	428 56.84
Total	606 100.00	147 100.00	753 100.00

```
. reg inlf Dkidslt6
```

Source	SS	df	MS
Model	7.89105463	1	7.89105463
Residual	176.836701	751	.23546831
Total	184.727756	752	.245648611

Number of obs	=	753
F(1, 751)	=	33.51
Prob > F	=	0.0000
R-squared	=	0.0427
Adj R-squared	=	0.0414
Root MSE	=	.48525

inlf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Dkidslt6	-.2582677	.0446138	-5.79	0.000	-.3458502	-.1706852
_cons	.6188119	.019712	31.39	0.000	.5801148	.657509

- [$\rightarrow$] Interpret parameters of the OLS model

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
    - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
  - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - Also, $E(y/x1, x2, ...xk) = P(y = 1/x1, x2, ...xk) * 1 + P(y = 0/x1, x2, ...xk) * 0 = P(y = 1/x1, x2, ...xk)$

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
  - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - Also, $E(y/x1, x2, ...xk) = P(y = 1/x1, x2, ...xk) * 1 + P(y = 0/x1, x2, ...xk) * 0 = P(y = 1/x1, x2, ...xk)$
  - Thus $P(y = 1/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
  - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - Also, $E(y/x1, x2, ...xk) = P(y = 1/x1, x2, ...xk) * 1 + P(y = 0/x1, x2, ...xk) * 0 = P(y = 1/x1, x2, ...xk)$
  - Thus $P(y = 1/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - And, $P(y = 0/x1, x2, ...xk) = 1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...))$

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
  - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - Also, $E(y/x1, x2, ...xk) = P(y = 1/x1, x2, ...xk) * 1 + P(y = 0/x1, x2, ...xk) * 0 = P(y = 1/x1, x2, ...xk)$
  - Thus $P(y = 1/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - And, $P(y = 0/x1, x2, ...xk) = 1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...))$
  - $\rightarrow$ Probability of `success' or offailure' are linear functions of the x_j (hence, the 'linear probability' model)

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
  - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - Also, $E(y/x1, x2, ...xk) = P(y = 1/x1, x2, ...xk) * 1 + P(y = 0/x1, x2, ...xk) * 0 = P(y = 1/x1, x2, ...xk)$
  - Thus $P(y = 1/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - And, $P(y = 0/x1, x2, ...xk) = 1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...))$
  - $\rightarrow$ Probability of `success' or offailure' are linear functions of the x_j (hence, the 'linear probability' model)
- β_1= ?

The LPM model: specification

- Let's consider the linear model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...) + u$
- We assume (A1)' to (A4)' are verified
- How interpret the β_j ?
  - $E(y/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - Also, $E(y/x1, x2, ...xk) = P(y = 1/x1, x2, ...xk) * 1 + P(y = 0/x1, x2, ...xk) * 0 = P(y = 1/x1, x2, ...xk)$
  - Thus $P(y = 1/x1, x2, ...xk) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...)$
  - And, $P(y = 0/x1, x2, ...xk) = 1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + (...))$
  - $\rightarrow$ Probability of `success' or offailure'` are linear functions of the x_j (hence, the 'linear probability' model)
- β_1= ?
  - $= \frac{\triangle\ P(y = 1/x1, x2, ...xk)}{\triangle\ x1}$: this is the change in the probability of 'success' following a change in x_1 (ceteris paribus)

The LPM model: coefficient interpretation (1)

inlf	Dkidslt6 0	1	Total
0	231 38.12	94 63.95	325 43.16
1	375 61.88	53 36.05	428 56.84
Total	606 100.00	147 100.00	753 100.00

```
. reg inlf Dkidslt6
```

Source	SS	df	MS
Model	7.89105463	1	7.89105463
Residual	176.836701	751	.23546831
Total	184.727756	752	.245648611

Number of obs	=	753
F(1, 751)	=	33.51
Prob > F	=	0.0000
R-squared	=	0.0427
Adj R-squared	=	0.0414
Root MSE	=	.48525

inlf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Dkidslt6	-.2582677	.0446138	-5.79	0.000	-.3458502	-.1706852
_cons	.6188119	.019712	31.39	0.000	.5801148	.657509

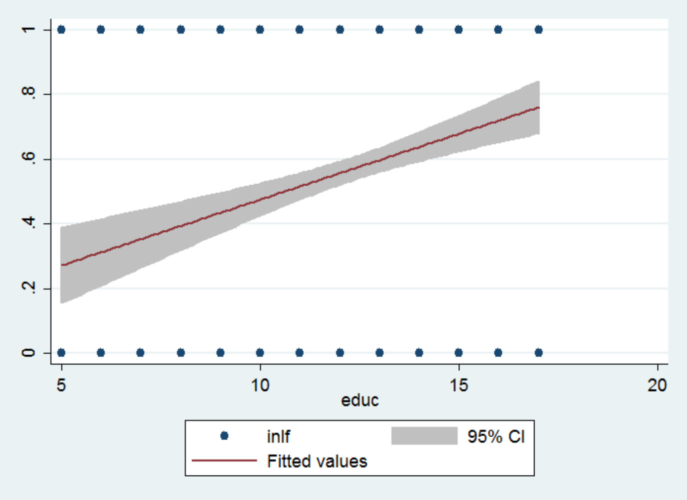
- [$\rightarrow$] Interpret parameters of the OLS model

The LPM model: coefficient interpretation (2)

```
. reg inlf educ

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(1, 751)       =     27.32
       Model |  6.48414537         1  6.48414537   Prob > F        =    0.0000
    Residual |   178.24361       751  .237341691   R-squared       =    0.0351
-------------+----------------------------------   Adj R-squared   =    0.0338
       Total |  184.727756       752  .245648611   Root MSE        =    .48718

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0407226   .0077911     5.23   0.000     .0254278    .0560175
       _cons |   .0680402     .09736     0.70   0.485    -.1230899    .2591703
------------------------------------------------------------------------------
```

- [$\rightarrow$] Interpret parameters of the OLS model

The LPM model: intuitions of limits



Limits of the LPM model: unplausible predicted probabilities

- From the graphic, it is clear that for some values of *educ*, we might find predicted probability above 1

Limits of the LPM model: unplausible predicted probabilities

- From the graphic, it is clear that for some values of *educ*, we might find predicted probability above 1
- Yet the probabilities should lie within the range (0,1).

Limits of the LPM model: unplausible predicted probabilities

- From the graphic, it is clear that for some values of *educ*, we might find predicted probability above 1
- Yet the probabilities should lie within the range (0,1).
- A solution is to truncate the probabilities at 0 or 1, so that a probability of, say, 1.2, would be set to 1.

Limits of the LPM model: unplausible predicted probabilities

- From the graphic, it is clear that for some values of *educ*, we might find predicted probability above 1
- Yet the probabilities should lie within the range (0,1).
- A solution is to truncate the probabilities at 0 or 1, so that a probability of, say, 1.2, would be set to 1.
- However, there are at least two reasons why this is still not adequate

Limits of the LPM model: unplausible predicted probabilities

- From the graphic, it is clear that for some values of *educ*, we might find predicted probability above 1
- Yet the probabilities should lie within the range (0,1).
- A solution is to truncate the probabilities at 0 or 1, so that a probability of, say, 1.2, would be set to 1.
- However, there are at least two reasons why this is still not adequate
    - risk that too many observations for which the estimated probabilities are exactly zero or one.

Limits of the LPM model: unplausible predicted probabilities

- From the graphic, it is clear that for some values of *educ*, we might find predicted probability above 1
- Yet the probabilities should lie within the range (0,1).
- A solution is to truncate the probabilities at 0 or 1, so that a probability of, say, 1.2, would be set to 1.
- However, there are at least two reasons why this is still not adequate
  - risk that too many observations for which the estimated probabilities are exactly zero or one.
  - how plausible is it to assume that a woman's probability to work is exactly 1 if she has a doctorate or close ?

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)
- Since the disturbance term changes systematically with the explanatory variables, the former will also be heteroscedastic

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)
- Since the disturbance term changes systematically with the explanatory variables, the former will also be heteroscedastic
  - $V(u/x1) = P(y = 1/x1) * [1 - \beta_0 - \beta_1 x1]^2 + P(y = 0/x1) * [-\beta_0 - \beta_1 x1]^2$

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)

- Since the disturbance term changes systematically with the explanatory variables, the former will also be heteroscedastic
  - $V(u/x1) = P(y = 1/x1) * [1 - \beta_0 - \beta_1 x1]^2 + P(y = 0/x1) * [-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]^2 + (1 - \beta_0 - \beta_1 x1)[-\beta_0 - \beta_1 x1]^2$

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)
- Since the disturbance term changes systematically with the explanatory variables, the former will also be heteroscedastic
  - $V(u/x1) = P(y = 1/x1) * [1 - \beta_0 - \beta_1 x1]^2 + P(y = 0/x1) * [-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]^2 + (1 - \beta_0 - \beta_1 x1)[-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]$

Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)

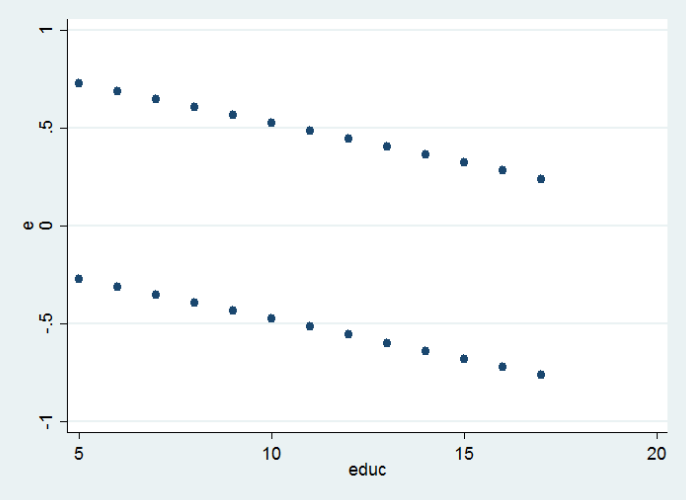
- Since the disturbance term changes systematically with the explanatory variables, the former will also be heteroscedastic
  - $V(u/x1) = P(y = 1/x1) * [1 - \beta_0 - \beta_1 x1]^2 + P(y = 0/x1) * [-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]^2 + (1 - \beta_0 - \beta_1 x1)[-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]$

- $\rightarrow \widehat{\beta_1}$ is unbiased but $V(\widehat{\beta_1})$ is biased

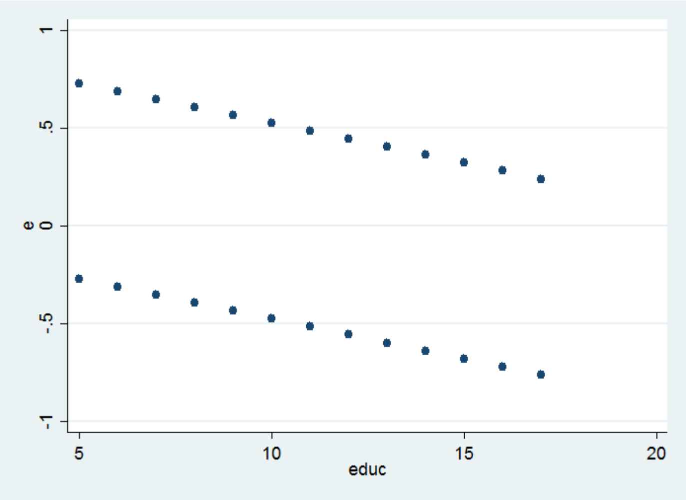
Limits of the LPM model: non-normality of errors and heteroscedasticity

- Since y only takes two values, for given x, the disturbance term will also only take on one of two values. Hence the error term cannot plausibly be assumed to be normally distributed
  - ok if n large (see chapter in Wooldridge on inferences when n is large)
- Since the disturbance term changes systematically with the explanatory variables, the former will also be heteroscedastic
  - $V(u/x1) = P(y = 1/x1) * [1 - \beta_0 - \beta_1 x1]^2 + P(y = 0/x1) * [-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]^2 + (1 - \beta_0 - \beta_1 x1)[-\beta_0 - \beta_1 x1]^2$
  - $V(u/x1) = (\beta_0 + \beta_1 x1)[1 - \beta_0 - \beta_1 x1]$
- $\rightarrow \widehat{\beta_1}$ is unbiased but $V(\widehat{\beta_1})$ is biased
- $\rightarrow$ We need to compute heteroscedasticity-robust standard errors (we know how to do; see previous chapter)

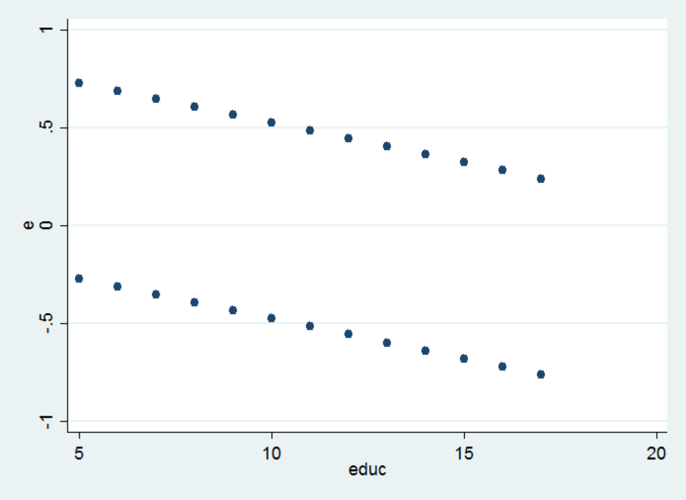
- predict y



- predict y
- gen e=inlf -y



- predict y
- gen e=inlf -y
- twoway (scatter e educ, sort)



Mathematical proof

- $y = xB + e$

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$
- or $E(y) = 1\textit{p(y=1)} + \textit{0}p(y=0) = p(y=1)$

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$
- or $E(y) = 1p(y{=}1) + 0p(y{=}0) = p(y{=}1)$
- donc $E(y|x) = xB = p(Y{=}1|x)$

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$
- or $E(y) = 1\mathit{p(y{=}1)} + \mathit{0}p(y{=}0) = p(y{=}1)$
- donc $E(y|x) = xB = p(Y{=}1|x)$
- d'un autre côté puisque Y est binaire

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$
- or $E(y) = 1\mathit{p(y{=}1)} + \mathit{0}p(y{=}0) = p(y{=}1)$
- donc $E(y|x) = xB = p(Y{=}1|x)$
- d'un autre côté puisque Y est binaire
- $V(Y) = p(1\text{-}p)$

Mathematical proof

- y = xB + e
- donc E(y|x) = xB car E(e|x) = 0
- or E(y) = 1*p(y=1)* + *0*p(y=0) = p(y=1)
- donc E(y|x) = xB = p(Y=1|x)
- d'un autre côté puisque Y est binaire
- V(Y) = p(1-p)
- vu que p=xB

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$
- or $E(y) = 1\mathit{p(y=1)} + \mathit{0}p(y=0) = p(y=1)$
- donc $E(y|x) = xB = p(Y=1|x)$
- d'un autre côté puisque Y est binaire
- $V(Y) = p(1\text{-}p)$
- vu que $p=xB$
- on peut aussi ecrire que

Mathematical proof

- y = xB + e
- donc E(y|x) = xB car E(e|x) = 0
- or E(y) = 1*p(y=1)* + *0*p(y=0) = p(y=1)
- donc E(y|x) = xB = p(Y=1|x)
- d'un autre côté puisque Y est binaire
- V(Y) = p(1-p)
- vu que p=xB
- on peut aussi ecrire que
- V(Y) = xB*(1-xB)

Mathematical proof

- $y = xB + e$
- donc $E(y|x) = xB$ car $E(e|x) = 0$
- or $E(y) = 1\,p(y{=}1) + 0\,p(y{=}0) = p(y{=}1)$
- donc $E(y|x) = xB = p(Y{=}1|x)$
- d'un autre côté puisque Y est binaire
- $V(Y) = p(1\text{-}p)$
- vu que $p{=}xB$
- on peut aussi ecrire que
- $V(Y) = xB^*(1\text{-}xB)$
- et varie donc avec avec X

- predict y

```
. tabstat e, by(educ) stat(sd)

Summary for variables: e
     by categories of: educ
```

educ	sd
5	.5
6	.5477226
7	.46291
8	.4982729
9	.5066228
10	.5052578
11	.5057805
12	.4974587
13	.4925448
14	.4826398
15	.5135526
16	.4689614
17	.3631584
Total	.4868532

- predict y
- gen e=inlf -y

```
. tabstat e, by(educ) stat(sd)

Summary for variables: e
     by categories of: educ
```

educ	sd
5	.5
6	.5477226
7	.46291
8	.4982729
9	.5066228
10	.5052578
11	.5057805
12	.4974587
13	.4925448
14	.4826398
15	.5135526
16	.4689614
17	.3631584
Total	.4868532

```
. reg inlf educ

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(1, 751)       =     27.32
       Model |  6.48414537         1  6.48414537   Prob > F        =    0.0000
    Residual |   178.24361       751  .237341691   R-squared       =    0.0351
-------------+----------------------------------   Adj R-squared   =    0.0338
       Total |  184.727756       752  .245648611   Root MSE        =    .48718

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0407226   .0077911     5.23   0.000     .0254278    .0560175
       _cons |   .0680402     .09736     0.70   0.485    -.1230899    .2591703
------------------------------------------------------------------------------

. reg inlf educ, ro

Linear regression                               Number of obs     =        753
                                                F(1, 751)         =      31.01
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0351
                                                Root MSE          =     .48718

------------------------------------------------------------------------------
             |               Robust
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0407226   .0073125     5.57   0.000     .0263673     .055078
       _cons |   .0680402   .0928069     0.73   0.464    -.1141516     .250232
------------------------------------------------------------------------------
```

Limits of the LPM model: conceptual ?

- Model the binary outcome or the process underlying the realization of the outcome ?

The probit and logit models

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$
- $y = 1$ if $y^* > 0$ and 0 otherwise

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$
- $y = 1$ if $y^* > 0$ and 0 otherwise
- y^* is not observed but y is

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$
- $y = 1$ if $y^* > 0$ and 0 otherwise
- y^* is not observed but y is
- Think of y^* as the net utility associated with a decision

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$
- $y = 1$ if $y^* > 0$ and 0 otherwise
- y^* is not observed but y is
- Think of y^* as the net utility associated with a decision
- $P(y = 1/x_1) =?$

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$
- $y = 1$ if $y^* > 0$ and 0 otherwise
- y^* is not observed but y is
- Think of y^* as the net utility associated with a decision
- $P(y = 1/x_1) = ?$
  - $= P(u > (-\beta_0 - \beta_1 x_1)/x_1) = ?$

Preliminary steps

- $y^* = \beta_0 + \beta_1 x_1 + u$
- $y = 1$ if $y^* > 0$ and 0 otherwise
- y^* is not observed but y is
- Think of y^* as the net utility associated with a decision
- $P(y = 1/x_1) =?$
  - $= P(u > (-\beta_0 - \beta_1 x_1)/x_1) =?$
  - It depends on the distribution of u

The logit and probit models: specifications

Two options in the litterature

- u follows a normal distribution

The logit and probit models: specifications

Two options in the litterature

- u follows a normal distribution
  - $P(u > (-\beta_0 - \beta_1 x_1)/x_1) = 1 - \Phi(-\beta_0 - \beta_1 x_1) = \Phi(\beta_0 + \beta_1 x_1)$ with Φ the cumulative normal distribution %which is symetric

The logit and probit models: specifications

Two options in the litterature

- u follows a normal distribution
  - $P(u > (-\beta_0 - \beta_1 x_1)/x_1) = 1 - \Phi(-\beta_0 - \beta_1 x_1) = \Phi(\beta_0 + \beta_1 x_1)$ with Φ the cumulative normal distribution %which is symetric
- u follows a logistic distribution

The logit and probit models: specifications

Two options in the litterature

- u follows a normal distribution
  - $P(u > (-\beta_0 - \beta_1 x_1)/x_1) = 1 - \Phi(-\beta_0 - \beta_1 x_1) = \Phi(\beta_0 + \beta_1 x_1)$ with Φ the cumulative normal distribution %which is symetric
- u follows a logistic distribution
  - $P(u > (-\beta_0 - \beta_1 x_1)/x_1) = 1 - \Lambda(-\beta_0 - \beta_1 x_1) = \Lambda(\beta_0 + \beta_1 x_1)$ with Λ the cumulative logistic distribution (which is symetric)

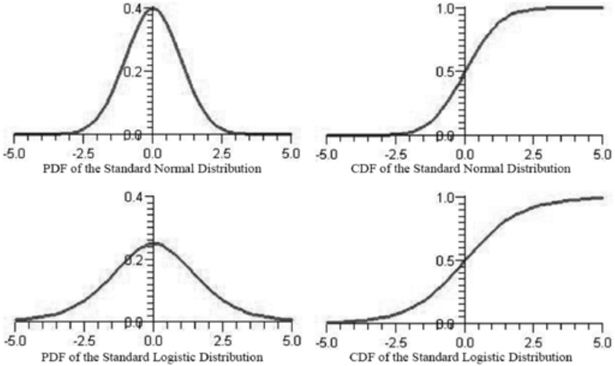


Figure 1. The Standard Normal and Standard Logistic Probability Distributions

Source : Park (2010)

The logit and probit models: specifications (2)

Probit model

- $P(y = 1/x_1) = P(u > (-\beta_0 - \beta_1 x_1)/x_1) = \Phi(\beta_0 + \beta_1 x_1)$

The logit and probit models: specifications (2)

Probit model

- $P(y = 1/x_1) = P(u > (-\beta_0 - \beta_1 x_1)/x_1) = \Phi(\beta_0 + \beta_1 x_1)$

The logit and probit models: specifications (2)

Probit model

- $P(y = 1/x_1) = P(u > (-\beta_0 - \beta_1 x_1)/x_1) = \Phi(\beta_0 + \beta_1 x_1)$

Logit model

- $P(y = 1/x_1) = P(u > (-\beta_0 - \beta_1 x_1)/x_1) = \Lambda(\beta_0 + \beta_1 x_1)$

The logit and probit models: specifications (3)

- Logit and probit models will not produce predicted probabilities above 1 (or below 0)

The logit and probit models: specifications (3)

- Logit and probit models will not produce predicted probabilities above 1 (or below 0)
- Logit and probit models are not linear (and cannot be made linear by a transformation) and thus are not estimable using OLS

The logit and probit models: specifications (3)

- Logit and probit models will not produce predicted probabilities above 1 (or below 0)
- Logit and probit models are not linear (and cannot be made linear by a transformation) and thus are not estimable using OLS
- Instead, maximum likelihood is usually used to estimate the parameters of the model

Estimation by maximum likelihood - Intuition

We seek β to 'maximize' the likelihood to observe our sample

- Let's write $G(\beta X)$ the cumulative distribution function (normal or logistic)

Estimation by maximum likelihood - Intuition

We seek β to 'maximize' the likelihood to observe our sample

- Let's write $G(\beta X)$ the cumulative distribution function (normal or logistic)
- We already know that the probability to observe $y_i = 1$ equals $G(x_i\beta)$ and the probability to observe $y_i = 0$ equals $1 - G(x_i\beta)$

Estimation by maximum likelihood - Intuition

We seek β to 'maximize' the likelihood to observe our sample

- Let's write $G(\beta X)$ the cumulative distribution function (normal or logistic)
- We already know that the probability to observe $y_i = 1$ equals $G(x_i\beta)$ and the probability to observe $y_i = 0$ equals $1 - G(x_i\beta)$
- We note $\ell_i(\beta)$ the likelihood to observe i $\ell_i(\beta) = G(X_i\beta)^{y_i} * [1 - G(x_i\beta)]^{1-y_i}$ with $y_i = (0;1)$ \(check that if y=1, $\ell_i(\beta) = G(X_i\beta)$ and that if y=0, $\ell_i(\beta) = 1 - G(X_i\beta)$)

Estimation by maximum likelihood - Intuition

We seek β to 'maximize' the likelihood to observe our sample

- Let's write $G(\beta X)$ the cumulative distribution function (normal or logistic)
- We already know that the probability to observe $y_i = 1$ equals $G(x_i\beta)$ and the probability to observe $y_i = 0$ equals $1 - G(x_i\beta)$
- We note $\ell_i(\beta)$ the likelihood to observe i $\ell_i(\beta) = G(X_i\beta)^{y_i} * [1 - G(x_i\beta)]^{1-y_i}$ with $y_i = (0;1)$ \(check that if y=1, $\ell_i(\beta) = G(X_i\beta)$ and that if y=0, $\ell_i(\beta) = 1 - G(X_i\beta)$)
- The log-likelihood to observe i is $Log[\ell_i(\beta)] = y_i Log[G(x_i\beta)] + (1 - y_i)Log[1 - G(x_i\beta)]$

Estimation by maximum likelihood - Intuition

We seek β to 'maximize' the likelihood to observe our sample

- Let's write $G(\beta X)$ the cumulative distribution function (normal or logistic)
- We already know that the probability to observe $y_i = 1$ equals $G(x_i\beta)$ and the probability to observe $y_i = 0$ equals $1 - G(x_i\beta)$
- We note $\ell_i(\beta)$ the likelihood to observe i $\ell_i(\beta) = G(X_i\beta)^{y_i} * [1 - G(x_i\beta)]^{1-y_i}$ with $y_i = (0;1)$ \(check that if y=1, $\ell_i(\beta) = G(X_i\beta)$ and that if y=0, $\ell_i(\beta) = 1 - G(X_i\beta)$)
- The log-likelihood to observe i is $Log[\ell_i(\beta)] = y_i Log[G(x_i\beta)] + (1 - y_i)Log[1 - G(x_i\beta)]$
- The log-likelihood to observe our entire sample is $\Sigma Log[\ell_i(\beta)] = \Sigma[y_i Log[G(x_i\beta)] + (1 - y_i)Log[1 - G(x_i\beta)]]$

Estimation by maximum likelihood - The probit and logit estimators

- We max $\Sigma Log[\ell_i(\beta)] = \Sigma[y_i Log[G(x_i\beta)] + (1 - y_i) Log[1 - G(x_i\beta)]]$ (solved by computers since FO conditions imply to solve for $k + 1$ equations)

Estimation by maximum likelihood - The probit and logit estimators

- We max $\Sigma Log[\ell_i(\beta)] = \Sigma[y_i Log[G(x_i\beta)] + (1-y_i)Log[1-G(x_i\beta)]]$ (solved by computers since FO conditions imply to solve for $k+1$ equations)
- The $\widehat{\beta}$ that maximize $\Sigma Log[\ell_i(\beta)]$ with G the normal cumulative distribution are named the probit estimator

Estimation by maximum likelihood - The probit and logit estimators

- We max $\Sigma Log[\ell_i(\beta)] = \Sigma[y_i Log[G(x_i\beta)] + (1 - y_i)Log[1 - G(x_i\beta)]]$ (solved by computers since FO conditions imply to solve for $k + 1$ equations)
- The $\widehat{\beta}$ that maximize $\Sigma Log[\ell_i(\beta)]$ with G the normal cumulative distribution are named the probit estimator
- The $\widehat{\beta}$ that maximize $\Sigma Log[\ell_i(\beta)]$ with G the logistic cumulative distribution are named the logit estimator

Estimation by maximum likelihood - Properties of the probit and logit estimators

The $\widehat{\beta}$ obtained by ML (under the hypothesis that u follows the assumed distribution)

- Are consistent

Estimation by maximum likelihood - Properties of the probit and logit estimators

The $\widehat{\beta}$ obtained by ML (under the hypothesis that u follows the assumed distribution)

- Are consistent
- If n large, are efficient

Estimation by maximum likelihood - Properties of the probit and logit estimators

The $\widehat{\beta}$ obtained by ML (under the hypothesis that u follows the assumed distribution)

- Are consistent
- If n large, are efficient
- If n large, are normally distributed

Estimation by maximum likelihood - Properties of the probit and logit estimators

The $\widehat{\beta}$ obtained by ML (under the hypothesis that u follows the assumed distribution)

- Are consistent
- If n large, are efficient
- If n large, are normally distributed
- If n large, inferences are made using same tools as in OLS models

Estimation by maximum likelihood - Properties of the probit and logit estimators

The $\widehat{\beta}$ obtained by ML (under the hypothesis that u follows the assumed distribution)

- Are consistent
- If n large, are efficient
- If n large, are normally distributed
- If n large, inferences are made using same tools as in OLS models
    - NB: Joint hypothesis tests are based on the Wald statistic which follows a Khi^2 distribution

Estimation by maximum likelihood - Application

```
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.06651
Iteration 2:   log likelihood = -401.30273
Iteration 3:   log likelihood = -401.30219
Iteration 4:   log likelihood = -401.30219

Probit regression                                 Number of obs     =        753
                                                  LR chi2(7)        =     227.14
                                                  Prob > chi2       =     0.0000
Log likelihood = -401.30219                       Pseudo R2         =     0.2206
```

inlf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0120237	.0048398	-2.48	0.013	-.0215096	-.0025378
educ	.1309047	.0252542	5.18	0.000	.0814074	.180402
exper	.1233476	.0187164	6.59	0.000	.0866641	.1600311
expersq	-.0018871	.0006	-3.15	0.002	-.003063	-.0007111
age	-.0528527	.0084772	-6.23	0.000	-.0694678	-.0362376
kidslt6	-.8683285	.1185223	-7.33	0.000	-1.100628	-.636029
kidsge6	.036005	.0434768	0.83	0.408	-.049208	.1212179
_cons	.2700768	.508593	0.53	0.595	-.7267472	1.266901

```
.
end of do-file
```

- How interpret the coefficient β_j on x_j ?

- How interpret the coefficient β_j on x_j ?
  - As we will see below, β_j is not a marginal effect ($\frac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$) and cannot be interpreted as such

- How interpret the coefficient β_j on x_j ?
  - As we will see below, β_j is not a marginal effect ($\frac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$) and cannot be interpreted as such
  - But β_j is a component of the marginal effect of x_j on the probability of success

- How interpret the coefficient β_j on x_j ?
  - As we will see below, β_j is not a marginal effect $(\frac{\Delta P(y_i = 1/x_i)}{\Delta x_j})$ and cannot be interpreted as such
  - But β_j is a component of the marginal effect of x_j on the probability of success
  - Precisely, β_j multiplied by a factor *that is always positive* gives the marginal effect of interest

- How interpret the coefficient β_j on x_j ?
  - As we will see below, β_j is not a marginal effect $(\frac{\Delta P(y_i = 1/x_i)}{\Delta x_j})$ and cannot be interpreted as such
  - But β_j is a component of the marginal effect of x_j on the probability of success
  - Precisely, β_j multiplied by a factor *that is always positive* gives the marginal effect of interest
  - So regarding β_j, the only thing we can interpret is its sign (but not is value since it does not measure a marginal effect)

Interpreting of probit/logit model results - Marginal effects at the mean

- Recall $P(y_i = 1/x_i) = \Phi(x_i\beta)$ (if a probit model)

Interpreting of probit/logit model results - Marginal effects at the mean

- Recall $P(y_i = 1/x_i) = \Phi(x_i\beta)$ (if a probit model)
- If x_j is continuous: $\frac{\Delta P(y_i = 1/x_i)}{\Delta x_j} = \beta_j \varphi(x_i\beta)$ with φ the derivative function of Φ and β_j the probit coefficient on x_j (indeed, [g(f(x)]'=f'g')

Interpreting of probit/logit model results - Marginal effects at the mean

- Recall $P(y_i = 1/x_i) = \Phi(x_i\beta)$ (if a probit model)
- If x_j is continuous: $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j} = \beta_j\varphi(x_i\beta)$ with φ the derivative function of Φ and β_j the probit coefficient on x_j (indeed, [g(f(x)]'=f'g')
  - The sign of $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$ is determined by the one of β_j (since φ $>$0)

Interpreting of probit/logit model results - Marginal effects at the mean

- Recall $P(y_i = 1/x_i) = \Phi(x_i\beta)$ (if a probit model)
- If x_j is continuous: $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j} = \beta_j \varphi(x_i\beta)$ with φ the derivative function of Φ and β_j the probit coefficient on x_j (indeed, [g(f(x)]'=f'g')
  - The sign of $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$ is determined by the one of β_j (since φ $>$0)
  - $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$ varies with the value of other x (usually we choose the sample mean)

Interpreting of probit/logit model results - Marginal effects at the mean

- Recall $P(y_i = 1/x_i) = \Phi(x_i\beta)$ (if a probit model)
- If x_j is continuous: $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j} = \beta_j \varphi(x_i\beta)$ with φ the derivative function of Φ and β_j the probit coefficient on x_j (indeed, [g(f(x)]'=f'g')
  - The sign of $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$ is determined by the one of β_j (since φ >0)
  - $\dfrac{\Delta P(y_i = 1/x_i)}{\Delta x_j}$ varies with the value of other x (usually we choose the sample mean)
- If x_2 is discrete : its change is associated with a change in $P(y_i = 1/x_i)$ of the following amount $\Phi(\beta_0 + \beta_1 x_1 + \beta_2 * 1 + (...) + \beta_k x_k) - \Phi(\beta_0 + \beta_1 x_1 + \beta_2 * 0 + (...) + \beta_k x_k)$

Interpreting of probit/logit model results - Marginal effects at the mean

- Recall $P(y_i = 1/x_i) = \Phi(x_i\beta)$ (if a probit model)
- If x_j is continuous: $\frac{\Delta P(y_i=1/x_i)}{\Delta x_j} = \beta_j \varphi(x_i\beta)$ with φ the derivative function of Φ and β_j the probit coefficient on x_j (indeed, [g(f(x)]'=f'g')
  - The sign of $\frac{\Delta P(y_i=1/x_i)}{\Delta x_j}$ is determined by the one of β_j (since φ >0)
  - $\frac{\Delta P(y_i=1/x_i)}{\Delta x_j}$ varies with the value of other x (usually we choose the sample mean)
- If x_2 is discrete : its change is associated with a change in $P(y_i = 1/x_i)$ of the following amount $\Phi(\beta_0 + \beta_1 x_1 + \beta_2 * 1 + (...) + \beta_k x_k) - \Phi(\beta_0 + \beta_1 x_1 + \beta_2 * 0 + (...) + \beta_k x_k)$
  - The marginal effect varies with the value of other x (usually we choose the sample mean)

Application: marginal effect of one unit change in education ?

```
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.06651
Iteration 2:   log likelihood = -401.30273
Iteration 3:   log likelihood = -401.30219
Iteration 4:   log likelihood = -401.30219

Probit regression                               Number of obs   =        753
                                                LR chi2(7)      =     227.14
                                                Prob > chi2     =     0.0000
Log likelihood = -401.30219                     Pseudo R2       =     0.2206
```

inlf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0120237	.0048398	-2.48	0.013	-.0215096	-.0025378
educ	.1309047	.0252542	5.18	0.000	.0814074	.180402
exper	.1233476	.0187164	6.59	0.000	.0866641	.1600311
expersq	-.0018871	.0006	-3.15	0.002	-.003063	-.0007111
age	-.0528527	.0084772	-6.23	0.000	-.0694678	-.0362376
kidslt6	-.8683285	.1185223	-7.33	0.000	-1.100628	-.636029
kidsge6	.036005	.0434768	0.83	0.408	-.049208	.1212179
_cons	.2700768	.508593	0.53	0.595	-.7267472	1.266901

```
. margins , dydx( * ) atmeans

Conditional marginal effects                    Number of obs     =         7
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper expersq age kidslt6 kidsge6
at           : nwifeinc        =    20.12896 (mean)
               educ            =    12.28685 (mean)
               exper           =    10.63081 (mean)
               expersq         =    178.0385 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interva
nwifeinc	-.0046962	.0018903	-2.48	0.013	-.0084012	-.00099
educ	.0511287	.0098592	5.19	0.000	.0318051	.07045
exper	.0481771	.0073278	6.57	0.000	.0338149	.06253
expersq	-.0007371	.0002347	-3.14	0.002	-.001197	-.00027
age	-.0206432	.0033079	-6.24	0.000	-.0271265	-.01415
kidslt6	-.3391514	.0463581	-7.32	0.000	-.4300117	-.24829
kidsge6	.0140628	.0169852	0.83	0.408	-.0192275	.04735

Application: marginal effect of one unit change in education ? (manual computation)

```
.
. *computation by hand
. estat summarize

  Estimation sample probit                 Number of obs =        753

  -------------------------------------------------------------------
      Variable |       Mean      Std. Dev.          Min          Max
  -------------+-----------------------------------------------------
          inlf |   .5683931      .4956295             0            1
      nwifeinc |   20.12896       11.6348     -.0290575           96
          educ |   12.28685      2.280246             5           17
         exper |   10.63081       8.06913             0           45
               |
       c.exper#|
       c.exper |   178.0385      249.6308             0         2025
               |
           age |   42.53785      8.072574            30           60
       kidslt6 |   .2377158       .523959             0            3
       kidsge6 |   1.353254      1.319874             0            8
  -------------------------------------------------------------------

. matrix list r(stats)

r(stats)[8,4]
                mean          sd         min         max
    inlf   .56839309   .49562951           0           1
nwifeinc   20.128964   11.634797   -.0290575          96
    educ   12.286853   2.2802458           5          17
   exper    10.63081   8.0691299           0          45
c.exper#
 c.exper   178.03851   249.63085           0        2025
     age   42.537849    8.072574          30          60
 kidslt6    .2377158   .52395904           0           3
 kidsge6   1.3532537   1.3198739           0           8

. matrix r = r(stats)

. scalar f1 = normalden(_b[nwifeinc]*r[2,1]+_b[educ]*r[3,1]+_b[exper]*r[4,1]+_b[c.exper#c.exper]*r[5,1]+_b[age]*r[6,1
> ] +_b[kidslt6]*r[7,1] + _b[kidsge6]*r[8,1]+ _b[_cons])
```

Application: marginal effect of one unit change in education ? \ (in logit)

```
. logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.38502
Iteration 2:   log likelihood = -401.76569
Iteration 3:   log likelihood = -401.76515
Iteration 4:   log likelihood = -401.76515

Logistic regression                             Number of obs     =        753
                                                LR chi2(7)        =     226.22
                                                Prob > chi2       =     0.0000
Log likelihood = -401.76515                     Pseudo R2         =     0.2197
```

inlf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0213452	.0084214	-2.53	0.011	-.0378509	-.0048394
educ	.2211704	.0434396	5.09	0.000	.1360303	.3063105
exper	.2058695	.0320569	6.42	0.000	.1430391	.2686999
expersq	-.0031541	.0010161	-3.10	0.002	-.0051456	-.0011626
age	-.0880244	.014573	-6.04	0.000	-.116587	-.0594618
kidslt6	-1.443354	.2035849	-7.09	0.000	-1.842373	-1.044335
kidsge6	.0601122	.0747897	0.80	0.422	-.086473	.2066974
_cons	.4254524	.8603697	0.49	0.621	-1.260841	2.111746

```
. margins , dydx( * ) atmeans

Conditional marginal effects                    Number of obs     =          7
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper expersq age kidslt6 kidsge6
at           : nwifeinc        =    20.12896 (mean)
               educ            =    12.28685 (mean)
               exper           =    10.63081 (mean)
               expersq         =    178.0385 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interva
nwifeinc	-.0051901	.0020482	-2.53	0.011	-.0092045	-.00117
educ	.0537773	.0105608	5.09	0.000	.0330785	.07447
exper	.0500569	.0078247	6.40	0.000	.0347209	.0653
expersq	-.0007669	.0002477	-3.10	0.002	-.0012524	-.00028
age	-.021403	.0035398	-6.05	0.000	-.0283408	-.01446
kidslt6	-.3509498	.0496395	-7.07	0.000	-.4482414	-.25365
kidsge6	.0146162	.0181884	0.80	0.422	-.0210324	.05026

Application: marginal effect of one unit change in education ? \ (in LPM)

```
. reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6
```

Source	SS	df	MS			
Model	48.8080578	7	6.97257968	Number of obs	=	753
Residual	135.919698	745	.182442547	F(7, 745)	=	38.22
				Prob > F	=	0.0000
				R-squared	=	0.2642
				Adj R-squared	=	0.2573
Total	184.727756	752	.245648611	Root MSE	=	.42713

inlf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nwifeinc	-.0034052	.0014485	-2.35	0.019	-.0062488	-.0005616
educ	.0379953	.007376	5.15	0.000	.023515	.0524756
exper	.0394924	.0056727	6.96	0.000	.0283561	.0506287
expersq	-.0005963	.0001848	-3.23	0.001	-.0009591	-.0002335
age	-.0160908	.0024847	-6.48	0.000	-.0209686	-.011213
kidslt6	-.2618105	.0335058	-7.81	0.000	-.3275875	-.1960335
kidsge6	.0130122	.013196	0.99	0.324	-.0128935	.0389179
_cons	.5855192	.154178	3.80	0.000	.2828442	.8881943

Partial effects and slope estimate comparison

- In probit: marginal effect of education, at the mean of other variables equals 5.11 (coeff or slope estimate equals 0.13)

Partial effects and slope estimate comparison

- In probit: marginal effect of education, at the mean of other variables equals 5.11 (coeff or slope estimate equals 0.13)
- In logit: marginal effect of education, at the mean of other variables equals 5.37 (coeff or slope estimate equals 0.22)

Partial effects and slope estimate comparison

- In probit: marginal effect of education, at the mean of other variables equals 5.11 (coeff or slope estimate equals 0.13)
- In logit: marginal effect of education, at the mean of other variables equals 5.37 (coeff or slope estimate equals 0.22)
  - To make the logit and probit slope roughly estimates comparable, we can either multiply the probit estimates by 1.6 (for instance, $0.13*1,6 = 0.21 \sim 0.22$), or multiply the logit estimates by .625

Partial effects and slope estimate comparison

- In probit: marginal effect of education, at the mean of other variables equals 5.11 (coeff or slope estimate equals 0.13)
- In logit: marginal effect of education, at the mean of other variables equals 5.37 (coeff or slope estimate equals 0.22)
  - To make the logit and probit slope roughly estimates comparable, we can either multiply the probit estimates by 1.6 (for instance, 0.13*1,6 $= 0.21 \sim 0.22$), or multiply the logit estimates by .625
- In LPM: marginal effect of education, ceteris paribus equals 3.79

Partial effects and slope estimate comparison

- In probit: marginal effect of education, at the mean of other variables equals 5.11 (coeff or slope estimate equals 0.13)
- In logit: marginal effect of education, at the mean of other variables equals 5.37 (coeff or slope estimate equals 0.22)
  - To make the logit and probit slope roughly estimates comparable, we can either multiply the probit estimates by 1.6 (for instance, 0.13*1,6 $=$ 0.21 $\sim$ 0.22), or multiply the logit estimates by .625
- In LPM: marginal effect of education, ceteris paribus equals 3.79
  - We can divide the logit slope estimates by 4 and the probit slope estimates by 2.5 to make them roughly comparable to the LPM estimates (for instance, 0.13/2,5 $=$ 0,052 $\sim$ 3.8 $\ldots$)

Fundamental difference between probit (logit) coefficient and lpm coefficient

- LPM: constant marginal effect (3.79 for education)

Fundamental difference between probit (logit) coefficient and lpm coefficient

- LPM: constant marginal effect (3.79 for education)
- Probit (or logit): non-constant marginal effect

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education $= 0.405$

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education $= 0.405$
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education $= 0.405$
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)
  - Predicted effect of having 9 years of education compared to 8 on the same outcome: $0.405 + 0.0379$

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education $= 0.405$
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)
  - Predicted effect of having 9 years of education compared to 8 on the same outcome: $0.405 + 0.0379$
  - Predicted difference in labor market participation between two average individuals except one has 16 years of education and the other has 15: $+ 0.0379$ too (since constant marginal effect)

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education = 0.405
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)
  - Predicted effect of having 9 years of education compared to 8 on the same outcome: $0.405 + 0.0379$
  - Predicted difference in labor market participation between two average individuals except one has 16 years of education and the other has 15: $+\ 0.0379$ too (since constant marginal effect)
- Probit : predicted probability to participate to the labor force with 8 years of education = 0.365. This result comes from

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education = 0.405
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)
  - Predicted effect of having 9 years of education compared to 8 on the same outcome: 0.405 + 0.0379
  - Predicted difference in labor market participation between two average individuals except one has 16 years of education and the other has 15: + 0.0379 too (since constant marginal effect)

- Probit : predicted probability to participate to the labor force with 8 years of education = 0.365. This result comes from
  - first, computing $\Phi(\beta_0 + \beta_1 * x_1 + \beta_2 * \mathbf{0} + (...) + \beta_k x_k)$ with x_2 measuring years of education (mean values for other variables)

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education = 0.405
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)
  - Predicted effect of having 9 years of education compared to 8 on the same outcome: 0.405 + 0.0379
  - Predicted difference in labor market participation between two average individuals except one has 16 years of education and the other has 15: + 0.0379 too (since constant marginal effect)
- Probit : predicted probability to participate to the labor force with 8 years of education = 0.365. This result comes from
  - first, computing $\Phi(\beta_0 + \beta_1 * x_1 + \beta_2 * \mathbf{0} + (...) + \beta_k x_k)$ with x_2 measuring years of education (mean values for other variables)
  - second,computing $\Phi(\beta_0 + \beta_1 * x_1 + \beta_2 * \mathbf{8} + (...) + \beta_k x_k) - \Phi(\beta_0 + \beta_1 x_1 + \beta_2 * \mathbf{0} + (...) + \beta_k x_k)$

Fundamental difference between probit (logit) coefficient and lpm coefficient (2)

- LPM: predicted probability to participate to the labor force with 8 years of education = 0.405
  - This result comes from replacing the variable years of education by 8 in the LPM equation (for other variables, we take the mean)
  - Predicted effect of having 9 years of education compared to 8 on the same outcome: 0.405 + 0.0379
  - Predicted difference in labor market participation between two average individuals except one has 16 years of education and the other has 15: + 0.0379 too (since constant marginal effect)
- Probit : predicted probability to participate to the labor force with 8 years of education = 0.365. This result comes from
  - first, computing $\Phi(\beta_0 + \beta_1 * x_1 + \beta_2 * \mathbf{0} + (...) + \beta_k x_k)$ with x_2 measuring years of education (mean values for other variables)
  - second,computing $\Phi(\beta_0 + \beta_1 * x_1 + \beta_2 * \mathbf{8} + (...) + \beta_k x_k) - \Phi(\beta_0 + \beta_1 x_1 + \beta_2 * \mathbf{0} + (...) + \beta_k x_k)$
  - third, summing the two values

Fundamental difference between probit (logit) coefficient and lpm coefficient (3)

- Probit: Predicted difference in labor market participation between two average individuals one with 9 years of education and one with 8 years of education?

Fundamental difference between probit (logit) coefficient and lpm coefficient (3)

- Probit: Predicted difference in labor market participation between two average individuals one with 9 years of education and one with 8 years of education?
    - we compute $\Phi(\beta_0+\beta_1*x_1+\beta_2*\mathbf{9}+(...)+\beta_k x_k)-\Phi(\beta_0+\beta_1 x_1+\beta_2*\mathbf{8}+(...)+\beta_k x_k)=$ **0.050**

Fundamental difference between probit (logit) coefficient and lpm coefficient (3)

- Probit: Predicted difference in labor market participation between two average individuals one with 9 years of education and one with 8 years of education?
  - we compute $\Phi(\beta_0+\beta_1*x_1+\beta_2*\mathbf{9}+(...)+\beta_k x_k)-\Phi(\beta_0+\beta_1 x_1+\beta_2*\mathbf{8}+(...)+\beta_k x_k) =$ **0.050**
- Probit: Predicted difference in labor market participation between two average individuals one with 16 years of education and one with 15 years of education?

Fundamental difference between probit (logit) coefficient and lpm coefficient (3)

- Probit: Predicted difference in labor market participation between two average individuals one with 9 years of education and one with 8 years of education?
  - we compute $\Phi(\beta_0+\beta_1*x_1+\beta_2*\mathbf{9}+(...)+\beta_k x_k)-\Phi(\beta_0+\beta_1 x_1+\beta_2*\mathbf{8}+(...)+\beta_k x_k) = \mathbf{0.050}$
- Probit: Predicted difference in labor market participation between two average individuals one with 16 years of education and one with 15 years of education?
- we compute $\Phi(\beta_0 + \beta_1 * x_1 + \beta_2 * \mathbf{16} + (...) + \beta_k x_k) - \Phi(\beta_0 + \beta_1 x_1 + \beta_2 * \mathbf{15} + (...) + \beta_k x_k) = \mathbf{0.043}\}$

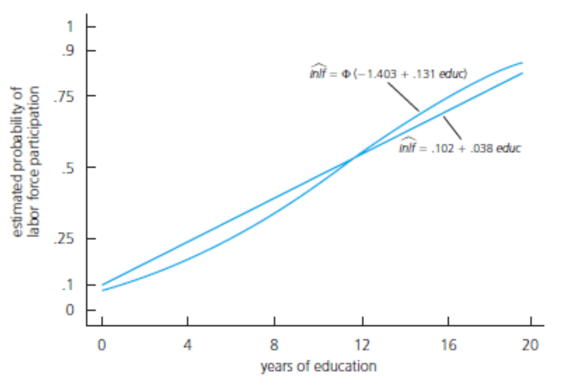
Fundamental difference between probit (logit) coefficient and lpm coefficient (4)

Computation for each margin . . . }

```
. local xb0 = _b[nwifeinc]*r[2,1]+_b[educ]*0+_b[exper]*r[4,1]+_b[expersq]*r[5,1]+_b[age]*r[6,1] +_b[kidslt6]*r[7,1] + _b[kidsge6]*r
> [8,1]+ _b[_cons]

. display normal(`xb0')
.08037296

. display normal(`xb0'+1*_b[educ]) - normal(`xb0')
.02137356

. display normal(`xb0'+2*_b[educ]) - normal(`xb0'+1*_b[educ])
.02523884

. display normal(`xb0'+3*_b[educ]) - normal(`xb0'+2*_b[educ])
.02929748

. display normal(`xb0'+4*_b[educ]) - normal(`xb0'+3*_b[educ])
.0334318

. display normal(`xb0'+5*_b[educ]) - normal(`xb0'+4*_b[educ])
.03750228

. display normal(`xb0'+6*_b[educ]) - normal(`xb0'+5*_b[educ])
.04135464

. display normal(`xb0'+7*_b[educ]) - normal(`xb0'+6*_b[educ])
.04482902

. display normal(`xb0'+8*_b[educ]) - normal(`xb0'+7*_b[educ])
.04777083

. display normal(`xb0'+9*_b[educ]) - normal(`xb0'+8*_b[educ])
.05004202

. display normal(`xb0'+10*_b[educ]) - normal(`xb0'+9*_b[educ])
.05153182

. display normal(`xb0'+11*_b[educ]) - normal(`xb0'+10*_b[educ])
.05216565

. display normal(`xb0'+12*_b[educ]) - normal(`xb0'+11*_b[educ])
.05191134

. display normal(`xb0'+13*_b[educ]) - normal(`xb0'+12*_b[educ])
.05078184

. display normal(`xb0'+14*_b[educ]) - normal(`xb0'+13*_b[educ])
.0488341
```

Fundamental difference between probit (logit) coefficient and lpm coefficient (5)

FIGURE 17.2 Estimated response probabilities with respect to education for the linear probability and probit models.



MLE and OLS estimator

Remarks

- Note that to estimate β using the OLS method, we do not need u to follow any distribution (we need u to follow a normal distribution for inference)

MLE and OLS estimator

Remarks

- Note that to estimate β using the OLS method, we do not need u to follow any distribution (we need u to follow a normal distribution for inference)
- To obtain β using the ML method, we need u to follow either a normal distribution or a logistic (if this is not the case, then we are not sure what the β measure)

Final remark (1): marginal effect of one unit change in experience ?

Be careful ! experience is entered in a non-linear way. To compute its marginal effect, we need to re-write the model !

```
. probit inlf nwifeinc educ exper c.exper#c.exper age kidslt6 kidsge6

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.06651
Iteration 2:   log likelihood = -401.30273
Iteration 3:   log likelihood = -401.30219
Iteration 4:   log likelihood = -401.30219

Probit regression                               Number of obs     =        753
                                                LR chi2(7)        =     227.14
                                                Prob > chi2       =     0.0000
Log likelihood = -401.30219                     Pseudo R2         =     0.2206
```

inlf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-.0120237	.0048398	-2.48	0.013	-.0215096	-.0025378
educ	.1309047	.0252542	5.18	0.000	.0814074	.180402
exper	.1233476	.0187164	6.59	0.000	.0866641	.1600311
c.exper#c.exper	-.0018871	.0006	-3.15	0.002	-.003063	-.0007111
age	-.0528527	.0084772	-6.23	0.000	-.0694678	-.0362376
kidslt6	-.8683285	.1185223	-7.33	0.000	-1.100628	-.636029
kidsge6	.036005	.0434768	0.83	0.408	-.049208	.1212179
_cons	.2700768	.508593	0.53	0.595	-.7267472	1.266901

```
. margins , dydx( * ) atmeans

Conditional marginal effects                    Number of obs     =
Model VCE    : OIM

Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc educ exper age kidslt6 kidsge6
at           : nwifeinc        =    20.12896 (mean)
               educ            =    12.28685 (mean)
               exper           =    10.63081 (mean)
               age             =    42.53785 (mean)
               kidslt6         =    .2377158 (mean)
               kidsge6         =    1.353254 (mean)
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Inte
nwifeinc	-.0045448	.0018286	-2.49	0.013	-.0081288	-.00
educ	.0494796	.0095876	5.16	0.000	.0306883	.06
exper	.0314576	.0031229	10.07	0.000	.0253368	.03
age	-.0199773	.0032404	-6.17	0.000	-.0263284	-.01
kidslt6	-.3282122	.0452473	-7.25	0.000	-.4168953	-.2
kidsge6	.0136092	.016439	0.83	0.408	-.0186107	.04

Goodness of fit - Predicted outcomes versus realized outcomes

- predict phat2

```
. tab p inlf, cell
```

Key
frequency   *cell percentage*

p	inlf 0	1	Total
0	205 27.22	80 10.62	285 37.85
1	120 15.94	348 46.22	468 62.15
Total	325 43.16	428 56.84	753 100.00

Goodness of fit - Predicted outcomes versus realized outcomes

- predict phat2
- gen p=phat2>0.5

```
. tab p inlf, cell
```

Key
frequency   *cell percentage*

p	inlf 0	1	Total
0	205 27.22	80 10.62	285 37.85
1	120 15.94	348 46.22	468 62.15
Total	325 [illegible]	428 [illegible]	753 [illegible]

Goodness of fit - Predicted outcomes versus realized outcomes

- predict phat2
- gen p=phat2>0.5
- tab p inlf, cell

```
. tab p inlf, cell
```

Key
frequency
cell percentage

p	inlf 0	inlf 1	Total
0	205 27.22	80 10.62	285 37.85
1	120 15.94	348 46.22	468 62.15
Total	325	428	753

Goodness of fit - The pseudo R2

- $= 1 - \frac{\Sigma Log[\ell_i(\beta)]_{uc}}{\Sigma Log[\ell_i(\beta)]_c}$

Goodness of fit - The pseudo R2

- $= 1 - \frac{\Sigma Log[\ell_i(\beta)]_{uc}}{\Sigma Log[\ell_i(\beta)]_c}$
- Note that if the model has no explanatory power, then $\Sigma Log[\ell_i(\beta)]_{uc} = \Sigma Log[\ell_i(\beta)]_c$ and the pseudo R2 $= 0$

Goodness of fit - The pseudo R2

- $= 1 - \frac{\Sigma Log[\ell_i(\beta)]_{uc}}{\Sigma Log[\ell_i(\beta)]_c}$
- Note that if the model has no explanatory power, then $\Sigma Log[\ell_i(\beta)]_{uc} = \Sigma Log[\ell_i(\beta)]_c$ and the pseudo R2 $= 0$
- In contrast if the model does very well – predict 1 for all observations with $y = 1$ and predict 0 for all observations with $y = 0$ –, then the log likelihood of the unrestricted model will approach 0 and the pseudo-R2 the unit)

Application: Determinants of private school enrolment in India

$\rightarrow$ TD