Extracting effects from non-linear models

Maarten L. Buis

Institut für Soziologie
Eberhard Karls Universität Tübingen
www.maartenbuis.nl

Introduction

- What is the effect of x on y?
  - Which effect do I choose: average marginal effects or marginal effects for someone with average values for the predictors or odds ratios, or...?
- Is the effect of x on y in group a the same as the effect of x on y in group b?
  - How to interpret interaction effects: marginal effect for interaction effects or ratio of odds ratios?
  - Is such a comparison of effects across groups even identified?
- How much of the effect of x on y can I explain with variable z?
  - How can I get indirect/mediator effects?

This introduction

- Quick review of
  - What is an effect?
  - What variables should we control for?
  - What is a non-linear model?

What is an effect?

- Almost always a comparison of means.
- Say we have data on the income of a number of males and a number of *comparable* females.
- The comparison of the mean income of males and females gives us the effect of gender on income.
- This comparison can take the form of a difference: women earn on average *x* euros/yen/pounds/dollars less than men,
- or it can take the form of a ratio: women earn on average *y*% less than men.

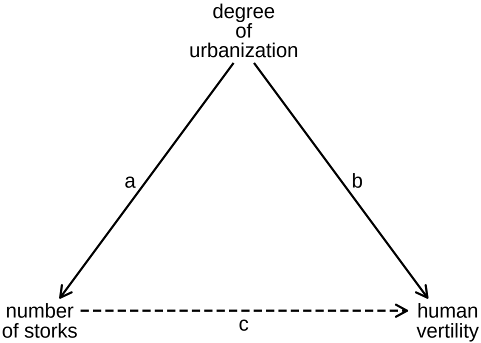
OK, but what about continuous variables?

- Say we want to know the effect of age on income.
- Still a comparison of groups, each 1 year apart.
- Easiest solution is to constrain all these effects to be the same.
- The default for “difference effects” in linear regression.
- The default for “ratio effects” in non-linear regression with the log link-function.

What variables do we need to control for?

- Confounding variables

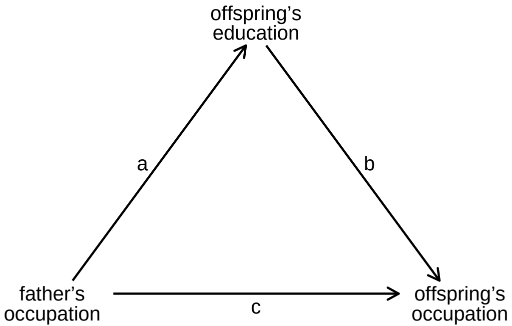
What variables do we need to control for?



What variables do we need to control for?

- Confounding variables
- Not intervening variables

What variables do we need to control for?



What variables do we need to control for?

- Confounding variables
- Not intervening variables
- Not idiosyncratic error/random noise/‘luck’
  - Many non-linear models exist to model a probability, an odds, a rate, or a hazard rate.
  - These concepts are defined by what we consider to be idiosyncratic error/random noise/‘luck’.
  - In these models the dependent variable is defined by what we choose not to control for.

Non-linear models

- $f(\text{E}(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$
- $f()$ is the link function, e.g.
  - logit link: $\log(\frac{u}{1-u})$
  - probit link: $\Phi(u)$
  - log link: $log(u)$
- an important characteristic of non-linear functions is that $f(\text{E}[y]) \neq \text{E}[f(y)]$
- Many non-linear models exist to accommodate
  - known bounds in the dependent variable, e.g. probability [0,1], odds, rate, hazard rate ≥ 0.
  - effects in terms of ratios

Adjusted predictions (1)

- Say we want to know what the effect of having a college-degree on the probability of never being married, while controlling for age and whether or not the respondent lives in the South of USA.
- We do a logitstic regression:

```
sysuse nlsw88, clear
logit union collgrad age south
```

- An effect is a comparison of means, so why not get a predicted probability for a typical respondent with a college-degree and without a college-degree?

Adjusted predictions (2)

- We could fix `age` and `south` at the mean and than predict the probability for the two groups:

Adjusted predictions (2)

```
. margin , at(collgrad=(0 1) (mean) south age) noatlegend

Adjusted predictions                              Number of obs   =       2246
Model VCE    : OIM

Expression   : Pr(never_married), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0874183   .0068627    12.74   0.000     .0739676     .100869
          2  |    .149139   .0154553     9.65   0.000     .1188472    .1794308
------------------------------------------------------------------------------
```

Adjusted predictions (2)

- We could fix `age` and `south` at the mean and than predict the probability for the two groups:
- Alternatively, We could predict the probabilities for all individuals, and than compute the mean probabilities within each group:

Adjusted predictions (2)

```
. margin , at(collgrad=(0 1)) noatlegend
Predictive margins                          Number of obs   =       2246
Model VCE     : OIM
Expression    : Pr(never_married), predict()
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0893587   .0068779    12.99   0.000     .0758783     .102839
          2  |   .1517616   .0154767     9.81   0.000     .1214279    .1820954
------------------------------------------------------------------------------
```

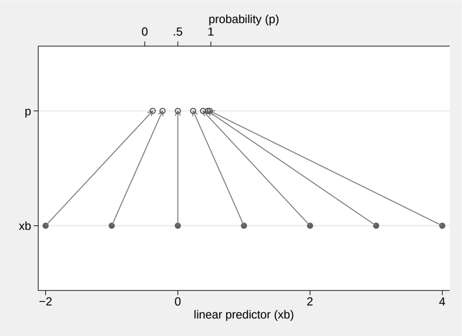
Adjusted predictions (2)

- We could fix `age` and `south` at the mean and than predict the probability for the two groups:
- Alternatively, We could predict the probabilities for all individuals, and than compute the mean probabilities within each group:
- Why are these not the same?

Adjusted predictions (3)

- The predicted probability is $\Lambda(xb)$, where
  - $\Lambda(u) = \frac{exp(u)}{1+exp(u)}$
  - $xb = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots$
- The first method consists of computing $\Lambda(\text{E}[xb])$.
- The second method consists of computing $\text{E}[\Lambda(xb)]$.

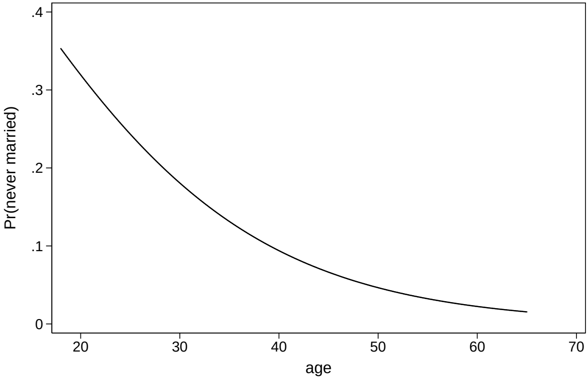
Adjusted predictions (3)



OK, but what about the effect of continuous variables?

- We want to summarize by how much the probability of being unmarried decreases when one gets a year older.
- This is a rate of change, or first derivative.
- In this context often called marginal effect.
- Problem: the relationship between age and the probability is non-linear, so there are many marginal effects

Effect of age



Effect of age (2)

```
. margins, dydx(age)  noatlegend                         ///
>           at((mean) collgrad south age=(35 45))

Conditional marginal effects                      Number of obs   =       2246
Model VCE    : OIM

Expression   : Pr(never_married), predict()
dy/dx w.r.t. : age

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
age          |
         _at |
          1  |  -.0086154   .0032458    -2.65   0.008     -.014977   -.0022538
          2  |  -.0046781   .0008927    -5.24   0.000    -.0064277   -.0029285
------------------------------------------------------------------------------
```

OK, but what about the effect of continuous variables?

- We want to summarize by how much the probability of being unmarried decreases when one gets a year older.
- This is a rate of change, or first derivative.
- In this context often called marginal effect.
- Problem: the relationship between age and the probability is non-linear, so there are many marginal effects
- Problem: We get different effects when first fix the explanatory variables and than compute the marginal effect or first compute the marginal effects for each individual and than average.

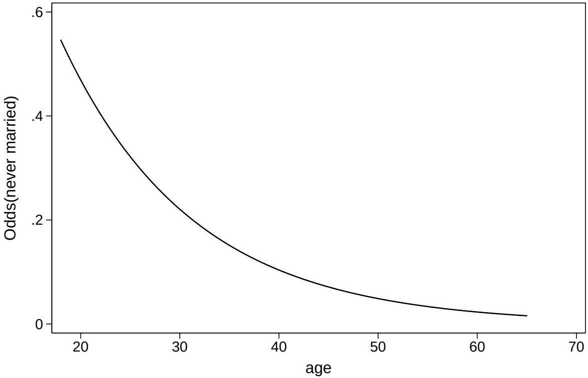
Effect of age (3)

```
. margins, dydx(age)  noatlegend                         ///
>           at((mean) collgrad south age=(35 45))

Conditional marginal effects                      Number of obs   =       2246
Model VCE    : OIM

Expression   : Pr(never_married), predict()
dy/dx w.r.t. : age

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
age          |
         _at |
          1  |  -.0086154   .0032458    -2.65   0.008     -.014977   -.0022538
          2  |  -.0046781   .0008927    -5.24   0.000    -.0064277   -.0029285
------------------------------------------------------------------------------
```

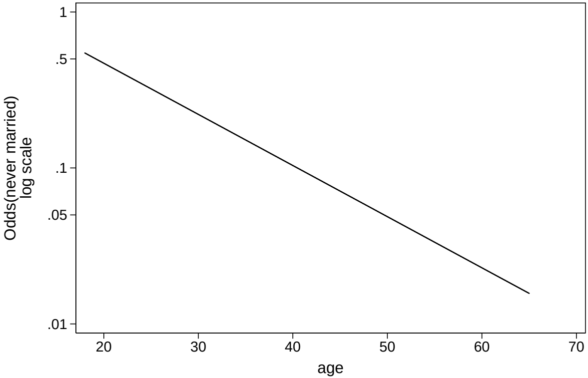
Too many effects

- So, what is the “true” effect?
- In a strick sense none of them, but they are all valid approximations
- There is an alternative that is not an approximation when the link function contains a logarithm.
- In that case the effect in terms of ratios is assumed to be constant.

Effect of age (4)



Effect of age (5)



Effect of age (3)

```
. gen byte baseline = 1

. gen c_age = age - 30

. logit never_married south c_age collgrad baseline, nolog nocons or

Logistic regression                               Number of obs   =       2246
                                                  Wald chi2(4)    =     940.91
Log likelihood = -736.65888                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
never_marr~d | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |   .8552155   .1219781    -1.10   0.273     .6466511    1.131048
       c_age |   .9272802   .0217551    -3.22   0.001     .8856065     .970915
    collgrad |   1.829794    .270653     4.08   0.000     1.369295    2.445159
    baseline |   .2041465   .0461335    -7.03   0.000     .1310948    .3179058
------------------------------------------------------------------------------
```

Marginal effects of interaction effects

- An interaction between two variables is included by creating a new variable that is the product of the two.
- In linear regression we can interpret the multiplicative term as how much the effect of variable 1 changes for a unit change in variable 2 (and vice versa).
- Ai and Norton (2003) pointed out that this does not work for marginal effects in non-linear models.
- The aim is find out how much the effect of x_1 changes for a unit change in x_2
- i.e. the cross partial derivative with respect to x_1 and x_2.
- These can be computed in Stata by `inteff` and `inteff3`.

Ratio effect of interaction effects

- The easier solution is to interpret interaction effects in terms of ratio effects.
- The interaction effect can now be interpreted as the ratio by which the effect of x_2 changes for a unit change in x_1

Interaction effects

```
. sysuse nlsw88, clear
(NLSW, 1988 extract)

. gen byte high_occ = occupation < 3 if occupation < .
(9 missing values generated)

. gen byte black = race == 2 if race < .

. drop if race == 3
(26 observations deleted)

. gen byte baseline = 1

. logit high_occ black##collgrad baseline, or nocons nolog

Logistic regression                               Number of obs   =       2211
                                                  Wald chi2(4)    =     504.62
Log likelihood = -1199.4399                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
    high_occ | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.black |   .4194072   .0655069    -5.56   0.000     .3088072    .5696188
  1.collgrad |   2.465411    .293568     7.58   0.000     1.952238    3.113478
             |
      black# |
    collgrad |
        1 1  |   1.479715   .4132536     1.40   0.161     .8559637    2.558003
             |
    baseline |   .3220524   .0215596   -16.93   0.000     .2824512    .3672059
------------------------------------------------------------------------------
```

Interaction effects

```
. margins , over(black collgrad) expression(exp(xb())) post

Predictive margins                              Number of obs   =       2211
Model VCE    : OIM

Expression   : exp(xb())
over         : black collgrad

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      black#|
    collgrad |
        0 0  |   .3220524   .0215596    14.94   0.000     .2797964    .3643084
        0 1  |   .7939914    .078188    10.15   0.000     .6407457    .9472371
        1 0  |   .1350711   .0190606     7.09   0.000      .097713    .1724292
        1 1  |   .4927536   .1032487     4.77   0.000       .29039    .6951173
------------------------------------------------------------------------------
```

Interaction effects

```
. lincom 0.black#1.collgrad - 0.black#0.collgrad
 ( 1)  - 0bn.black#0bn.collgrad + 0bn.black#1.collgrad = 0
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
(1)	.471939	.081106	5.82	0.000	.3129742	.6309038

```
. lincom 1.black#1.collgrad - 1.black#0.collgrad
 ( 1)  - 1.black#0bn.collgrad + 1.black#1.collgrad = 0
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
(1)	.3576825	.1049933	3.41	0.001	.1518994	.5634656

Latent variable interpretation of logistic regression

- Assume that there is some latent propensity of success y^*
- Someone gets a successes if $y^* > 0$ otherwise a failure.
- $y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$
- the scale of y^* is fixed by fixing the standard deviation of ε to a fixed number $\frac{\pi}{\sqrt{3}}$.
- If we compare effects across groups or models we have to assume that the residual variance is equal otherwise the scale of the dependent variable will differ.

Scenarios

- One way to get an idea about the size of this problem is to estimate various scenarios.
- The idea is that the heteroscedasticity comes from a (composite) unobserved variable, and to make assumptions regarding the size of the effect of this variable, its distribution, and how the effect changes when the observed variable of interest changes.
- The effect of the observed variables can than be estimated by integrating the likelihood function over this unobserved variable, which can be done by maximum simulated likelihood.
- This is implemented in Stata in `scenreg`.

Probability and odds interpretation of logistic regression

- The problem is that the scale of y^* is not defined
- We can solve that by interpreting the effects in terms of probabilities or odds, as these have a known scale.
- This does not do away with all arbitrariness:
  - the probability is defined in terms of what variables we chose to designate idiosyncratic error/luck
  - i.e. which variables we choose not to control for.
- Comparison of groups (interaction effects) can be solved this way, but comparisons of models with different explanatory variables (indirect effects) are still problematic.

Problem with naïve method

```
. drop _all

. set obs 60000
obs was 0, now 60000

. gen z = ceil(_n / 20000) - 1

. bys z: gen x = ceil(_n  / 10000) - 1

. tab x z

           |                    z
         x |         0          1          2 |     Total
-----------+---------------------------------+----------
         0 |    10,000     10,000     10,000 |    30,000
         1 |    10,000     10,000     10,000 |    30,000
-----------+---------------------------------+----------
     Total |    20,000     20,000     20,000 |    60,000

. set seed 12345

. gen y = runiform() < invlogit(-4 + 4*x + 2*z)
```

Problem with naïve method

```
. qui logit y x z
. est store direct
. local direct = _b[x]
.
. qui logit y x
. est store total
. local total _b[x]
.
. est tab direct total

----------------------------------------
    Variable |   direct       total     
-------------+--------------------------
           x |  4.0391332    2.6256242  
           z |   2.026339               
       _cons | -4.0452305   -1.3123133  
----------------------------------------

. di "naive indirect effect = " `total´ - `direct´
naive indirect effect = -1.413509
```

`ldecomp` solution:

Say we want to find the indirect effect of college education through occupation on union membership.

- Estimate a logistic regression with all variables.
- Predict the log odds for each respondent and transform these to proportions.
- Compute the average proportion for college-graduates and non-college-graduates, and transform back to log odds: the difference between these is the total effect.
- Compute the average proportion for college graduates, assuming they have the distribution of occupation of the non-college-graduates.
- The only difference between the college graduates and the counterfactual group is the distribution of occupation, so this difference represents the indirect effect.
- The distribution of occupation remains constant when comparing the counterfactual group with the non-college graduates, so this difference represents the

Example

```
. sysuse nlsw88, clear
(NLSW, 1988 extract)

. gen byte high = occupation < 3 if occupation <.
(9 missing values generated)

. gen byte middle = occupation >= 3 & occupation < 7 if occupation < .
(9 missing values generated)

. ldecomp union south, direct(collgrad) indirect(high middle) at(south 0) or
(running _ldecomp on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Bootstrap results                               Number of obs      =      1869
                                                Replications       =        50

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1/0          |
       total |   1.657501   .1867359     4.49   0.000     1.329096     2.06705
   indirect1 |   .8958344   .0491377    -2.01   0.045     .8045225    .9975101
     direct1 |   1.850231   .2249036     5.06   0.000     1.458004    2.347974
   indirect2 |   .8872166   .0493551    -2.15   0.031     .7955693    .9894213
     direct2 |   1.868203   .2311406     5.05   0.000     1.465921    2.380881
------------------------------------------------------------------------------

in equation i/j (comparing groups i and j)
let the fist subscript of Odds be the distribution of the the indirect variable
let the second subscript of Odds be the conditional probabilities
Method 1: Indirect effect = Odds_ij/Odds_jj
            Direct effect = Odds_ii/Odds_ij
Method 2: Indirect effect = Odds_ii/Odds_ji
            Direct effect = Odds_ji/Odds_jj

value labels
   0 not college grad
   1 college grad
```

Conclusion

- There are two things that make non-linear models more difficult than non-linear models
  - The dependent variable is related to the independent variables via a non-linear function
  - The dependent variable is not directly observed, but a function of our model
- Often we can prevent the problem by using “ratio effects” instead of “difference effects”
- Sometimes we can bypass this problem by using a linear model as an approximation.
- Sometimes we will just have to use more complicated methods (`ldecomp`, `scenreg`, `inteff`, `inteff3`)