Discrete choice models in STATA

Francesco Mattioli

20612 – Political Science – Module I (Topics in Comparative Politics)
M.Sc. Politics and Policy Analysis

November 6, 2023

Introduction

- Many political and economic outcomes of interest can be measured through quantitative variables

- As many common outcomes have a qualitative, or categorical, structure:
  - Choosing between voting or not
  - Choosing which party to vote
  - Reporting own interest in politics

- If we want to analyze quantitatively these outcomes, basic regression models are not appropriate

Outline

Binary choice models

- We want to understand how individual characteristics, $\boldsymbol{x}$ (e.g. gender, age, income), predict electoral participation, y
- We use microdata from Round 10 of the European Social Survey, which asks *'Did you vote in the last national election?'*
- Two possible, mutually exclusive answers: *'No'* (0) or *Yes* (1)
- Voting is a binary random variable $Y \sim Bernoulli\,(p)$

$$y_i = \begin{cases} 1 & \textit{with prob. } p \\ 0 & \textit{with prob. } 1 - p \end{cases}$$

Modelling binary outcomes

- How to evaluate the impact of $\boldsymbol{x}$ on y?
- We could regress y on $\boldsymbol{x}$ to estimate the coefficients $\boldsymbol{\beta}$
- Individual's i probability of voting is then a function of his/her characteristics, to be modeled through any proper functional form $F(\cdot)$:

$$p_i \equiv Pr(y_i = 1|\boldsymbol{x}) = F(\boldsymbol{x}_i'\boldsymbol{\beta})$$

- Three popular forms for $F(\cdot)$ are available:

Model	Functional form	Probability p	Marginal effect of x_j
LPM	Linear function: $f(\cdot)$	$f(\boldsymbol{x}'\boldsymbol{\beta}) = \boldsymbol{x}'\boldsymbol{\beta}$	β_j
Logit	Logistic cdf: $\Lambda(\cdot)$	$\Lambda(\boldsymbol{x}'\boldsymbol{\beta}) = \frac{e^{\boldsymbol{x}'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}'\boldsymbol{\beta}}}$	$\Lambda(\boldsymbol{x}'\boldsymbol{\beta})\{1-\Lambda(\boldsymbol{x}'\boldsymbol{\beta})\}\beta_j$
Probit	Std. normal cdf: $\Phi(\cdot)$	$\Phi(\boldsymbol{x}'\boldsymbol{\beta}) = \int_{-\infty}^{\boldsymbol{x}'\boldsymbol{\beta}} \phi(z)\, dz$	$\phi(\boldsymbol{x}'\boldsymbol{\beta})\beta_j$

An example in STATA (code fully commented on BBoard)

```
// access the European Social Survey data portal at https://ess-search.nsd.no
// download the dataset "ESS10 - integrated file, edition 3.1" in STATA format (.dta) after registering to the website
// store it into a proper location on your laptop
clear all
cd "/Users/francescomattioli/Library/CloudStorage/OneDrive-UniversitaCommercialeLuigiBocconi/PhD/TA/20612 - Political Science/stata"
use "ESS10/ESS10.dta"
// We want to study the socio-demographic determinants of voter participation among Italians
// We are interested in variable "vote"
codebook vote        // Is vote a suitable binary variable?
recode vote (2 = 0) (3/.z = .), generate(turnout)
label variable turnout "Turnout (binary)"
label define turnout_labels 0 "No" 1 "Yes"
label values turnout turnout_labels
```

```
// Let's choose some covariates of interest, e.g. age, gender, education, and income
// Clean them in the same way, but more quickly

clonevar age = agea if agea < .                        // age
clonevar gender = gndr                                 // gender (no need to clean it)
clonevar educ_years = eduyrs    if eduyrs <= 30        // years of education (few very high values - outliers?)
clonevar income_d = hinctnta if hinctnta < .           // deciles of household net income

// Let's focus on Italy
keep if cntry=="IT"

// Finding variables of interest:
// - read the codebook provided with the dataset
// - explore variables using STATA's data screening commands
// - type keywords in the variable list (on the right of STATA's interface) and explore the variables retained
```

```
. summarize i.turnout age i.gender educ_years i.income_d, vsquish

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
     turnout |
         No  |      2,366     .239645     .426957          0          1
        Yes  |      2,366     .760355     .426957          0          1
         age |      2,597    51.58568    18.68979         15         90
      gender |
       Male  |      2,640        .475    .4994692          0          1
     Female  |      2,640        .525    .4994692          0          1
-------------+---------------------------------------------------------
  educ_years |      2,547    12.43659    4.239423          0         28
    income_d |
J - 1st d..  |      1,627    .0547019    .2274674          0          1
R - 2nd d..  |      1,627    .1290719    .3353826          0          1
C - 3rd d..  |      1,627    .1567302     .363658          0          1
M - 4th d..  |      1,627    .1352182    .3420616          0          1
-------------+---------------------------------------------------------
F - 5th d..  |      1,627    .1155501    .3197829          0          1
S - 6th d..  |      1,627    .1180086    .3227175          0          1
K - 7th d..  |      1,627     .122311    .3277454          0          1
P - 8th d..  |      1,627    .0823602    .2749972          0          1
D - 9th d..  |      1,627    .0590043    .2357052          0          1
-------------+---------------------------------------------------------
H - 10th ..  |      1,627    .0270436    .1622605          0          1
```

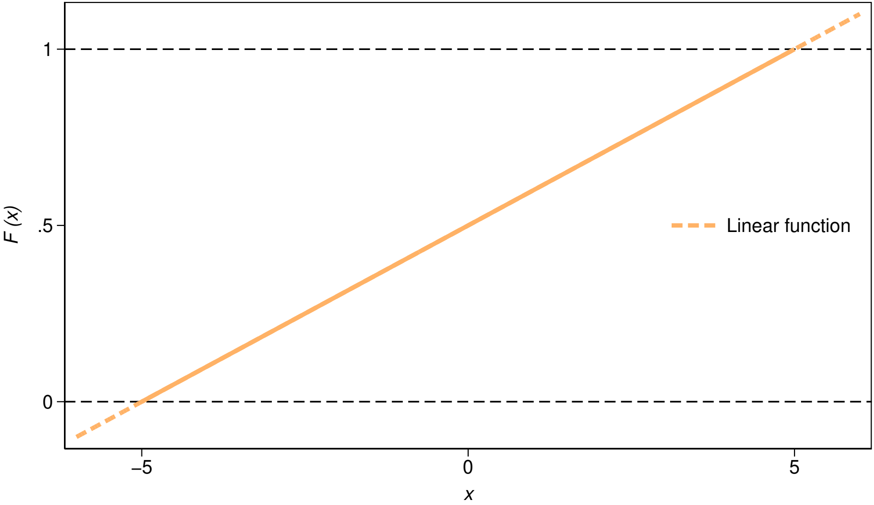
Linear Probability Model (LPM)

- Probability of voting is modelled as a linear function of $\boldsymbol{x}$, and estimated by Ordinary Least Squares (OLS):

$$Pr\left(y_i = 1|\boldsymbol{x}\right) = \boldsymbol{x}_i'\boldsymbol{\beta}$$

- PROs:
  - OLS estimation is quick and straightforward
  - Intuitive and direct interpretation of marginal effects
  - In practice very similar to non-linear models (as $n \to \infty$)
- CONs:
  - Predicted probabilities outside the unit interval ($p < 0$ or $p > 1$)
  - Standard errors are heteroskedastic ($Var\left(\varepsilon_i|\boldsymbol{x}_i\right) = \sigma_i^2$)
  - A one-unit increase in x_j changes y by $\widehat{\beta}_j$ regardless of the starting value of x_j (constant marginal effects): effects are estimated more (less) precisely near (away from) the center of the distribution of x_j

LPM: problems



```
. regress turnout age i.gender educ_years i.income_d

      Source |       SS           df       MS      Number of obs   =     1,461
-------------+----------------------------------   F(12, 1448)     =      8.02
       Model |  14.8504404        12  1.2375367   Prob > F        =    0.0000
    Residual |  223.547917     1,448  .154383921   R-squared       =    0.0623
-------------+----------------------------------   Adj R-squared   =    0.0545
       Total |  238.398357     1,460  .163286546   Root MSE        =    .39292

------------------------------------------------------------------------------
     turnout | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0022124   .0006678     3.31   0.001     .0009024    .0035224
             |
      gender |
     Female  |  -.0269749   .0207815    -1.30   0.194    -.0677401    .0137902
  educ_years |    .015086   .0027841     5.42   0.000     .0096246    .0205474
             |
    income_d |
R - 2nd decile  |   .0260417   .0522308     0.50   0.618    -.0764144    .1284978
C - 3rd decile  |   .1445538   .0509095     2.84   0.005     .0446895     .244418
M - 4th decile  |   .1588765   .0523927     3.03   0.002     .0561029    .2616501
F - 5th decile  |   .0939942    .053654     1.75   0.080    -.0112538    .1992421
S - 6th decile  |   .1288264   .0538023     2.39   0.017     .0232876    .2343652
K - 7th decile  |   .2251418   .0532091     4.23   0.000     .1207666    .3295171
P - 8th decile  |   .2104106   .0573682     3.67   0.000      .097877    .3229442
D - 9th decile  |   .1999662   .0610336     3.28   0.001     .0802425      .31969
H - 10th decile |   .1429715   .0782535     1.83   0.068    -.0105308    .2964737
             |
       _cons |   .3679212   .0729288     5.04   0.000     .2248638    .5109786
------------------------------------------------------------------------------
```

```
predict turnout_pr, xb                    // compute predicted probabilities and check their
      distribution
capture count if (turnout_pr < 0 | turnout_pr > 1) & turnout_pr != .        // count how many
      observations have a predicted probability outside the unit interval
display "LPM predictions outside unit interval: `r(N)'"

LPM predictions outside unit interval: 26

summarize turnout_pr

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  turnout_pr |      1,461    .7946612     .100854   .5230744   1.087262

estat hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of turnout

H0: Constant variance

chi2(1) =  79.20
Prob > chi2 = 0.0000
```

LPM: marginal effects

- In LPM they correspond to the estimated $\widehat{\beta}_{OLS}$ coefficients
- Interpretation: since the outcome is binary, $\widehat{\beta}_{OLS}$ is interpreted as the change in the probability of $y = 1$
- An example based on previous STATA output:
  - *Categorical variables*: individuals with income in the 5$^{\text{th}}$ decile have a non-significantly higher probability of voting than those in the 1$^{\text{st}}$ decile by ≈ 9 **percentage points** (pp) – **not** by 9%!
  - *Continuous variables*: an additional year of age significantly increases the probability of voting by ≈ 0.2 pp (regardless of the starting age)

```
margins i.varname               // predicted values of y for each catergory of
     varname (categorical)
margins, at(varname = #)        // predicted values of y at specified levels of
     varname (continuous)
margins, dydx(varlist)           // marginal effect of variables in varlist
```

Interactions and factor-variable notation

- The effect of x_j on y might differ depending on the level of x_k: this requires entering non-linear terms in the regression, i.e. interacting variables

- Interacting means multiplying two or more variables, thus generating additional coefficients that require careful interpretation

- STATA's factor-variable operators, combined with command **`margins`**, simplify this task

Operator	Description
`i.``varname`	indicators for each category of *varname*
`c.``varname`	*varname* treated as continuous
`varname`**`#`**`varname`	interaction of two variables
`varname`**`##`**`varname`	interacting variables **and** their interaction

- Three forms of interaction are possible

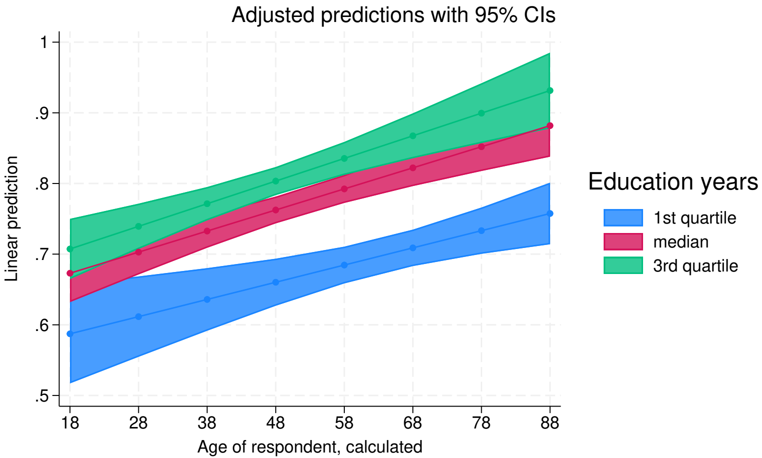
Interaction 1: continuous × continuous

```
. regress turnout c.age##c.educ_years, noheader
------------------------------------------------------------------------------------
           turnout | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------------+----------------------------------------------------------------
               age |   .0015524   .0015347     1.01   0.312    -.0014572     .004562
        educ_years |   .0151716   .0071382     2.13   0.034     .0011735    .0291697
                   |
c.age#c.educ_years |   .0001101   .0001173     0.94   0.348    -.0001199      .00034
                   |
             _cons |   .4220867   .0997573     4.23   0.000     .2264608    .6177126
------------------------------------------------------------------------------------
```

- $\widehat{\beta}_{\texttt{age}}$: one more year of age increases `turnout` by $\approx .16$ pp when `educ_years` is zero
- $\widehat{\beta}_{\texttt{educ\_years}}$: one more year of education increases `turnout` by ≈ 1.5 pp when `age` is zero
- $\widehat{\beta}_{\texttt{age\#educ\_years}}$: one more year of age increases `turnout` by $\approx .01$ pp for every additional year of `educ_years` (or viceversa)

Interaction 1: continuous × continuous

```
margins, at(age=(18(10)90) educ_years=(8 13 15))

marginsplot, recastci(rarea) legend(title("Education years") order(1 "1st
    quartile" 2 "median" 3 "3rd quartile"))
```



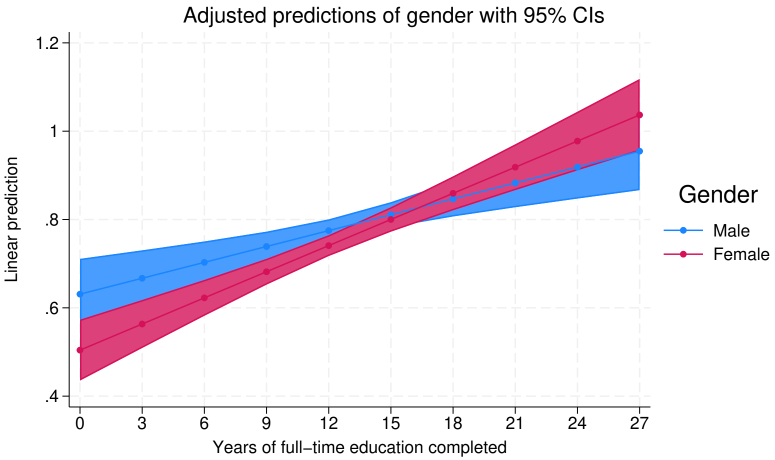
Interaction 2: continuous × categorical

```
. regress turnout i.gender##c.educ_years, noheader
------------------------------------------------------------------------------------
             turnout | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------------+---------------------------------------------------------------
              gender |
              Female |    -.12677    .0537889    -2.36   0.019    -.2322503   -.0212897
          educ_years |   .0119892    .0030309     3.96   0.000     .0060457    .0179327
                     |
gender#c.educ_years |
              Female |   .0077347    .0040556     1.91   0.057    -.0002183    .0156877
                     |
               _cons |   .6309944    .0407287    15.49   0.000     .5511254    .7108635
------------------------------------------------------------------------------------
```

- $\widehat{\beta}_{\texttt{Female}}$: women have a smaller `turnout` than men by $\approx .13$ pp when `educ_years` is zero
- $\widehat{\beta}_{\texttt{educ\_years}}$: one more year of education increases `turnout` by ≈ 1.2 pp for men
- $\widehat{\beta}_{\texttt{gender\#educ\_years}}$: women have a larger `turnout` than men by $\approx .8$ pp for every additional year of `educ_years`

Interaction 2: continuous × categorical

```
margins i.gender, at(educ_years=(0(3)28))

marginsplot, recastci(rarea) legend(title("Gender"))
```



Interaction 3: categorical × categorical

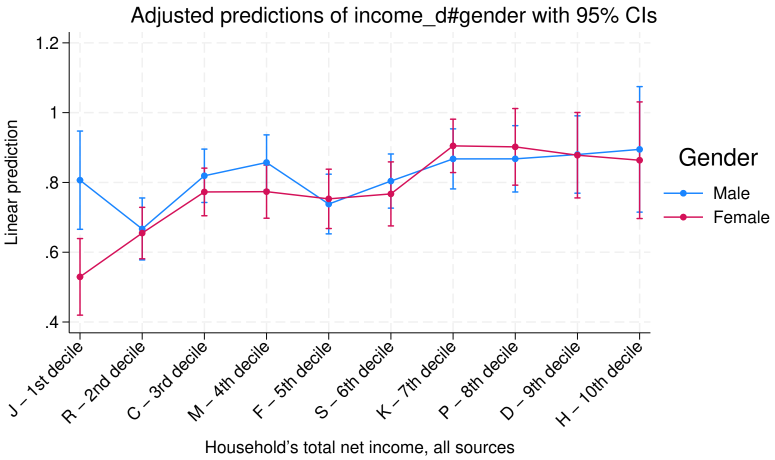
```
. regress turnout i.gender##i.income_d, noheader
------------------------------------------------------------------------------------------
                turnout | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
------------------------+-----------------------------------------------------------------
                 gender |
                Female  |  -.2770398   .0910069    -3.04   0.002    -.4555563   -.0985234
                        |
               income_d |
         R - 2nd decile |  -.1397849   .0848434    -1.65   0.100    -.3062114    .0266415
         C - 3rd decile |    .012596   .0816821     0.15   0.877    -.1476293    .1728214
         M - 4th decile |   .0506912   .0823444     0.62   0.538    -.1108332    .2122157
         F - 5th decile |  -.0683564   .0839772    -0.81   0.416    -.2330837     .096371
         S - 6th decile |    -.00253   .0819555    -0.03   0.975    -.1632916    .1582315
         K - 7th decile |   .0610183   .0841135     0.73   0.468    -.1039764    .2260129
         P - 8th decile |   .0611954   .0865995     0.71   0.480    -.1086757    .2310666
         D - 9th decile |   .0735484   .0913503     0.81   0.421    -.1056417    .2527385
        H - 10th decile |   .0882852   .1164288     0.76   0.448    -.1400982    .3166687
                        |
        gender#income_d |
  Female#R - 2nd decile |   .2652404   .1083635     2.45   0.014     .0526777    .4778032
  Female#C - 3rd decile |   .2307195   .1049419     2.20   0.028     .0248684    .4365706
  Female#M - 4th decile |   .1934819   .1068558     1.81   0.070    -.0161234    .4030872
  Female#F - 5th decile |   .2918858   .1098268     2.66   0.008     .0764527    .5073189
  Female#S - 6th decile |   .2402416   .1097054     2.19   0.029     .0250466    .4554365
  Female#K - 7th decile |   .3143319   .1082912     2.90   0.004     .1019109    .5267529
  Female#P - 8th decile |   .3113536   .1173101     2.65   0.008     .0812414    .5414658
  Female#D - 9th decile |   .2750886   .1239788     2.22   0.027     .0318954    .5182819
Female#H - 10th decile  |   .2459394   .1547424     1.59   0.112    -.0575989    .5494776
                        |
                  _cons |   .8064516   .0717715    11.24   0.000     .6656666    .9472366
------------------------------------------------------------------------------------------
```

Interaction 3: categorical $\times$ categorical

- $\widehat{\beta}_{\texttt{Female}}$: women have a smaller `turnout` than men by $\approx .28$ pp when `income_d` is zero
- $\widehat{\beta}_{\texttt{10th\_decile}}$: men in the 10^{th} income decile have a higher `turnout` by ≈ 9 pp than men in the 1^{st} decile (generalizes to other categories of `income_d`)
- $\widehat{\beta}_{\texttt{gender\#10th\_decile}}$: women in the 10^{th} income decile have a larger `turnout` than men in the same decile by $\approx .25$ pp compared to the difference between women and men in the 1^{st} decile (hint: think about a Diff-in-Diff coefficient; generalizes to other categories of `income_d`))

Interaction 3: categorical × categorical

```
margins i.income_d#i.gender

marginsplot, xlabel(, angle(45)) legend(title("Gender"))
```



Logit model and logistic regression

- To overcome the problems of LPM, the probability of voting can be modelled as a logistic function of $\boldsymbol{x}$:

$$p = Pr\left(y = 1|\boldsymbol{x}\right) = \Lambda\left(\boldsymbol{x'\beta}\right) = \frac{e^{\boldsymbol{x'\beta}}}{1 + e^{\boldsymbol{x'\beta}}}$$

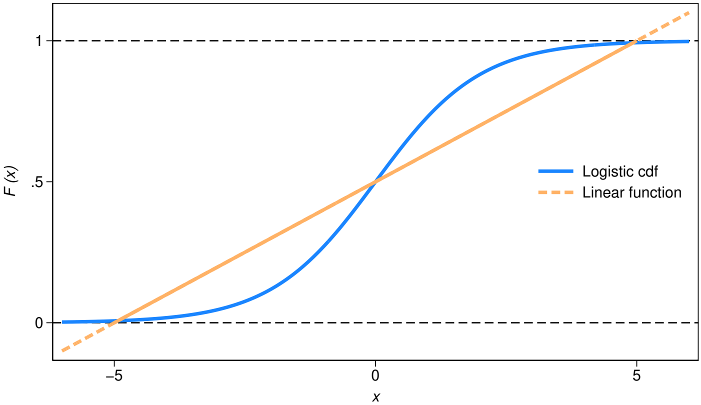
- The probability p of any binary outcome can alternatively be represented by means of the associated odds $\frac{p}{1-p}$

- The **log**istic un**it** model (hence logit) expresses the log-odds of the outcome as a linear function of $\boldsymbol{x}$ (easier to work with, being a log-linear model):

$$logit\left(p\right) = ln\left(\frac{p}{1-p}\right) = \boldsymbol{x'\beta}$$

- The two notations are equivalent and interchangeably referred to as logit model, the only difference lies in the interpretation of $\boldsymbol{\beta}$

Logit: solution to LPM problems



- Constrains probabilities between their natural boundaries $(0 \leq p \leq 1)$:

$$\lim_{x \to -\infty} Pr\left(y_i = 1 | \boldsymbol{x}\right) = 0 \qquad \lim_{x \to +\infty} Pr\left(y_i = 1 | \boldsymbol{x}\right) = 1$$

- Marginal effects differ with the point of evaluation $\boldsymbol{x}_i$

Logit: estimation

- Linear models are estimated by OLS: the estimated $\widehat{\beta}_{OLS}$ minimizes the residual sum of squares of the model (the coefficient that leads to the smallest error term)

- Non-linear models (also probit) are estimated by Maximum Likelihood (ML): the final $\widehat{\beta}_{MLE}$ maximizes the (log-)likelihood function (the coefficient that makes most likely the sample being analyzed)

- Computationally more demanding than OLS: starts with a guess $\widehat{\beta}_0$ and computes the likelihood, adjusts the initial guess and re-iterates the computation of the likelihood until it converges to the $\widehat{\beta}_{MLE}$ that makes the likelihood highest

- ML estimates are less reliable than OLS estimates in small samples

- Two ways to estimate a logit model in STATA

```
. logit turnout age i.gender educ_years i.income_d

Iteration 0:  Log likelihood = -741.77176
Iteration 1:  Log likelihood = -697.04399
Iteration 2:  Log likelihood = -694.84045
Iteration 3:  Log likelihood = -694.83169
Iteration 4:  Log likelihood = -694.83169

Logistic regression                                     Number of obs =  1,461
                                                        LR chi2(12)   =  93.88
                                                        Prob > chi2   = 0.0000
Log likelihood = -694.83169                             Pseudo R2     = 0.0633

------------------------------------------------------------------------------------
         turnout | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
             age |   .0162673   .0044331     3.67   0.000     .0075786    .0249561
                 |
          gender |
         Female  |  -.1801431   .1361288    -1.32   0.186    -.4469506    .0866644
      educ_years |   .1080394    .020023     5.40   0.000      .068795    .1472838
                 |
        income_d |
 R - 2nd decile  |   .1105034   .2810715     0.39   0.694    -.4403866    .6613935
 C - 3rd decile  |   .7045156   .2867744     2.46   0.014     .1424481    1.266583
 M - 4th decile  |   .8012036   .3029256     2.64   0.008     .2074803    1.394927
 F - 5th decile  |   .4087714   .2984958     1.37   0.171    -.1762696    .9938124
 S - 6th decile  |    .594014   .3083027     1.93   0.054    -.0102481    1.198276
 K - 7th decile  |   1.381398   .3417502     4.04   0.000     .7115801    2.051216
 P - 8th decile  |    1.22115   .3734159     3.27   0.001     .4892687    1.953032
 D - 9th decile  |   1.173607   .4058888     2.89   0.004     .3780799    1.969135
H - 10th decile  |   .7342881   .5498955     1.34   0.182    -.3434873    1.812064
                 |
           _cons |  -1.356762   .4661195    -2.91   0.004    -2.270339   -.4431842
------------------------------------------------------------------------------------
```

Logit: marginal effects

- In non-linear models $\widehat{\beta}_{MLE}$ coefficients do not correspond to marginal effects
- The marginal effect of x_j is the slope of a probability **curve** evaluated at specific values of $\boldsymbol{x}$ and $\widehat{\beta}$ (non-constant marginal effects):
  - The *sign* of the effect is the same as that of $\widehat{\beta}_j$
  - The *magnitude* of the effect is $\Lambda\left(\boldsymbol{x}'\widehat{\boldsymbol{\beta}}\right)\left\{1-\Lambda\left(\boldsymbol{x}'\widehat{\boldsymbol{\beta}}\right)\right\}\widehat{\beta}_j$
- Evaluations point must be chosen when reporting results from non-linear models
- It is possible to compare directly the relative effects of pairs of regressors (e.g. being in the 10$^{\text{th}}$ income decile corresponds to $.734/.108 \approx 7$ more years spent in eduction)

```
. logit turnout age i.gender educ_years i.income_d, or

Iteration 0:  Log likelihood = -741.77176
Iteration 1:  Log likelihood = -697.04399
Iteration 2:  Log likelihood = -694.84045
Iteration 3:  Log likelihood = -694.83169
Iteration 4:  Log likelihood = -694.83169

Logistic regression                                     Number of obs =  1,461
                                                        LR chi2(12)   =  93.88
                                                        Prob > chi2   = 0.0000
Log likelihood = -694.83169                             Pseudo R2     = 0.0633

------------------------------------------------------------------------------
     turnout | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |     1.0164   .0045058     3.67   0.000     1.007607     1.02527
             |
      gender |
     Female  |   .8351507    .113688    -1.32   0.186     .6395755    1.090531
  educ_years |   1.114092   .0223075     5.40   0.000     1.071217    1.158683
             |
    income_d |
 R - 2nd decile  |    1.11684    .313912     0.39   0.694     .6437875     1.93749
 C - 3rd decile  |   2.022867   .5801063     2.46   0.014     1.153093    3.548706
 M - 4th decile  |   2.228221   .6749853     2.64   0.008     1.230574     4.03468
 F - 5th decile  |   1.504968   .4492265     1.37   0.171     .8383919    2.701514
 S - 6th decile  |   1.811244   .5584114     1.93   0.054     .9898042    3.314399
 K - 7th decile  |   3.980463   1.360324     4.04   0.000     2.037208    7.777355
 P - 8th decile  |   3.391087   1.266286     3.27   0.001     1.631123    7.050032
 D - 9th decile  |   3.233637   1.312497     2.89   0.004      1.45948    7.164475
H - 10th decile  |   2.083998   1.145981     1.34   0.182     .7092925     6.12307
             |
       _cons |   .2574933   .1200226    -2.91   0.004     .1032772    .6419889
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Logit: marginal effects

- Alternatively, $\widehat{\beta}_j$ is interpreted as the effect of a one-unit change in x_j on $logit\,(p) = ln\left(\frac{p}{1-p}\right)$
- The transformation $e^{\widehat{\beta}_j}$ (also called odds-ratio) gives the **multiplicative** effect of a one-unit change in x_j on the odds $\frac{p}{1-p}$
  - $e^{\widehat{\beta}_j} > 1$ implies that the odds of voting are $e^{\widehat{\beta}_j}$ times larger; the odds of voting increase by $100\left(e^{\widehat{\beta}_j} - 1\right)\%$
  - $e^{\widehat{\beta}_j} < 1$ (but always > 0!) implies that the odds of voting decrease by a factor of $e^{\widehat{\beta}_j}$; the odds of voting decrease by $100\left(1 - e^{\widehat{\beta}_j}\right)\%$
- It is essential to know the starting values of the odds to quantify changes in probabilities correctly

Logit: marginal effects

- Because marginal effects depend on the point of evaluation, and if odds-ratios remain difficult to interpret, it is recommended to summarize marginal effects otherwise. Three common variants:
  - *Average marginal effect*: average of marginal effects for each individual
  - *Marginal effects at the mean*: marginal effects for the average individual (i.e. individual with average characteristics)
  - *Marginal effects at representative value*: marginal effects for a representative individual (i.e. individual with representative characteristics)
- To compute marginal effects of other kind of changes in x_j (not just one-unit changes) use `prchange` *`varlist`*
- It is possible to compute predicted probabilities as seen in the OLS case (`margins i.`*`varlist`* or `margins, at(c.`*`varlist`*` = #)`)

```
. margins, dydx(*)                     // average marginal effects of each variable

Average marginal effects                                        Number of obs = 1,461
Model VCE: OIM

Expression: Pr(turnout), predict()
dy/dx wrt:  age 2.gender educ_years 2.income_d 3.income_d 4.income_d 5.income_d 6.income_d 7.
     income_d 8.income_d 9.income_d 10.income_d

-------------------------------------------------------------------------------------
                |            Delta-method
                |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
----------------+--------------------------------------------------------------------
            age |   .0024823   .0006697     3.71   0.000     .0011697    .0037949
                |
         gender |
         Female |  -.0274486   .0206816    -1.33   0.184    -.0679838    .0130866
     educ_years |   .0164861   .0029881     5.52   0.000     .0106297    .0223426
                |
       income_d |
 R - 2nd decile |   .0229247   .0587337     0.39   0.696    -.0921914    .1380407
 C - 3rd decile |   .1293056   .0558197     2.32   0.021     .0199011    .2387102
 M - 4th decile |   .1437221   .0571827     2.51   0.012     .0316459    .2557982
 F - 5th decile |    .080091   .0597827     1.34   0.180    -.0370809    .1972629
 S - 6th decile |   .1118177   .0597302     1.87   0.061    -.0052514    .2288869
 K - 7th decile |   .2135951   .0557563     3.83   0.000     .1043149    .3228754
 P - 8th decile |   .1970155   .0594668     3.31   0.001     .0804627    .3135684
 D - 9th decile |   .1917146   .0627796     3.05   0.002     .0686688    .3147604
H - 10th decile |   .1338327   .0905325     1.48   0.139    -.0436079    .3112732
-------------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

```
. margins, dydx(*) atmeans noatlegend     // marginal effects of each variable for the average
     individual

Conditional marginal effects                           Number of obs = 1,461
Model VCE: OIM

Expression: Pr(turnout), predict()
dy/dx wrt:  age 2.gender educ_years 2.income_d 3.income_d 4.income_d 5.income_d 6.income_d 7.
     income_d 8.income_d 9.income_d 10.income_d

--------------------------------------------------------------------------------
                 |            Delta-method
                 |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-----------------+--------------------------------------------------------------
             age |   .0024602   .0006616     3.72   0.000     .0011635    .0037569
                 |
          gender |
          Female |  -.0271832   .0204741    -1.33   0.184    -.0673118    .0129453
      educ_years |   .0163395   .0029453     5.55   0.000     .0105667    .0221122
                 |
        income_d |
 R - 2nd decile  |   .0234455   .0601247     0.39   0.697    -.0943967    .1412877
 C - 3rd decile  |   .1305049   .0568114     2.30   0.022     .0191565    .2418533
 M - 4th decile  |   .1447572   .0580837     2.49   0.013     .0309152    .2585991
 F - 5th decile  |    .081365   .0609014     1.34   0.182    -.0379997    .2007296
 S - 6th decile  |   .1131272   .0606835     1.86   0.062    -.0058102    .2320646
 K - 7th decile  |   .2128163   .0565676     3.76   0.000     .1019458    .3236868
 P - 8th decile  |   .1968276   .0600304     3.28   0.001     .0791701    .3144851
 D - 9th decile  |   .1916939   .0631545     3.04   0.002     .0679134    .3154743
H - 10th decile  |   .1349877   .0905552     1.49   0.136    -.0424973    .3124726
--------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

```
. margins, dydx(*) at(age = 30 gender = 1 educ_years = 10 income_d = 4) noatlegend // marginal
      effects of each variable for a representative individual (e.g. 30-year-old woman with 10
      years of education and income in 4th decile)

Conditional marginal effects                            Number of obs = 1,461
Model VCE: OIM

Expression: Pr(turnout), predict()
dy/dx wrt:  age 2.gender educ_years 2.income_d 3.income_d 4.income_d 5.income_d 6.income_d 7.
      income_d 8.income_d 9.income_d 10.income_d

------------------------------------------------------------------------------------
                 |            Delta-method
                 |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
             age |   .0031793   .0010799     2.94   0.003     .0010627    .0052959
                 |
          gender |
          Female |  -.0366504   .0278018    -1.32   0.187     -.091141    .0178401
      educ_years |   .0211153   .0052713     4.01   0.000     .0107837     .031447
                 |
        income_d |
 R - 2nd decile  |   .0271331   .0691807     0.39   0.695    -.1084586    .1627249
 C - 3rd decile  |   .1615442   .0672301     2.40   0.016     .0297756    .2933128
 M - 4th decile  |   .1808622   .0695863     2.60   0.009     .0444757    .3172488
 F - 5th decile  |   .0975985   .0717145     1.36   0.174    -.0429593    .2381563
 S - 6th decile  |   .1384688    .072346     1.91   0.056    -.0033268    .2802644
 K - 7th decile  |   .2783284   .0695151     4.00   0.000     .1420812    .4145756
 P - 8th decile  |   .2546212   .0750535     3.39   0.001     .1075191    .4017233
 D - 9th decile  |   .2471179   .0800974     3.09   0.002     .0901299     .404106
H - 10th decile  |   .1675816   .1154887     1.45   0.147    -.0587721    .3939353
------------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

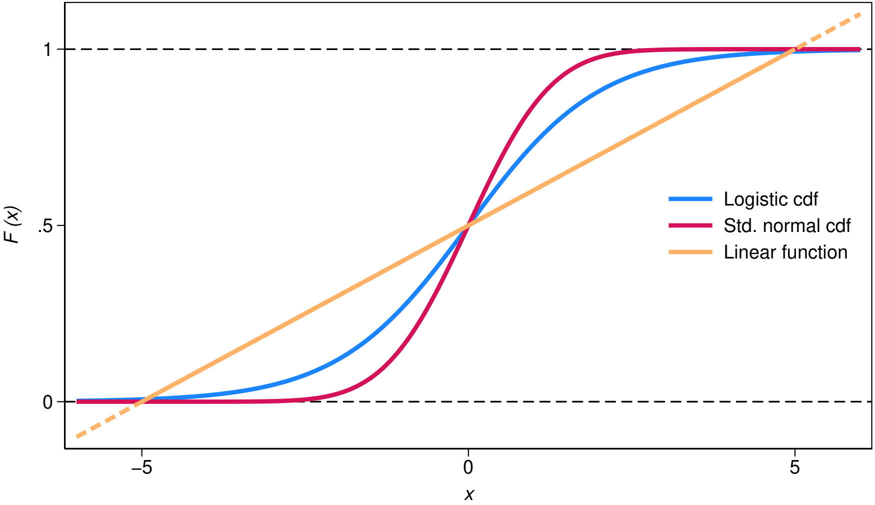
Probit

- Another popular method to model the probability of voting non-linearly is through a standard normal function of x:

$$p = Pr\left(y = 1|\boldsymbol{x}\right) = \Phi\left(\boldsymbol{x'\beta}\right) = \int_{-\infty}^{\boldsymbol{x'\beta}} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$$

- This is called **prob**ability un**it** model (hence probit)
- Similar to logit model:
  - Both densities are symmetric around the mean (bell-shaped)
  - Both models are estimated by maximum likelihood
  - Both have non-constant marginal effects

Probit: a logit with thinner tails



Probit: marginal effects

- The marginal effect of x_j depends on the specific values of $\boldsymbol{x}$ and $\widehat{\boldsymbol{\beta}}$:

$$\frac{1}{\sqrt{2\pi}} e^{\frac{-(\boldsymbol{x}'\widehat{\boldsymbol{\beta}})^2}{2}} \widehat{\beta}_j$$

- A one-unit change in x_j has an effect of $\widehat{\beta}_j$ on the z-score of y (measurement units of a standard normal distribution)
- Unlike the logit model, the probit model does not allow a transformation of variables that makes coefficients easier to interpret
- When reporting probit results it is recommended to resort on **`margins`** directly to compute interpretable predicted probabilities or marginal effects (STATA syntax seen for logit applies also for probit)

```
. probit turnout age i.gender educ_years i.income_d        // probit model estimation

Iteration 0:  Log likelihood = -741.77176
Iteration 1:  Log likelihood = -695.86802
Iteration 2:  Log likelihood = -695.31477
Iteration 3:  Log likelihood = -695.31474

Probit regression                                       Number of obs =  1,461
                                                        LR chi2(12)   =  92.91
                                                        Prob > chi2   = 0.0000
Log likelihood = -695.31474                             Pseudo R2     = 0.0626

---------------------------------------------------------------------------------
         turnout | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-----------------+---------------------------------------------------------------
             age |   .0089451   .0024794     3.61   0.000     .0040856    .0138046
                 |
          gender |
          Female |  -.1030726   .0778016    -1.32   0.185    -.2555608    .0494157
      educ_years |   .0595093     .01103     5.40   0.000     .0378908    .0811278
                 |
        income_d |
 R - 2nd decile  |   .0633083   .1730522     0.37   0.714    -.2758679    .4024845
 C - 3rd decile  |   .4109274   .1729947     2.38   0.018      .071864    .7499908
 M - 4th decile  |   .4650742   .1807062     2.57   0.010     .1108966    .8192518
 F - 5th decile  |   .2486845   .1814353     1.37   0.170    -.1069221    .6042912
 S - 6th decile  |   .3590502   .1850274     1.94   0.052    -.0035969    .7216973
 K - 7th decile  |   .7851377   .1946791     4.03   0.000     .4035736    1.166702
 P - 8th decile  |   .7105861   .2123559     3.35   0.001     .2943763    1.126796
 D - 9th decile  |   .6807951   .2299235     2.96   0.003     .2301533    1.131437
H - 10th decile  |   .4433589   .3046471     1.46   0.146    -.1537384    1.040456
                 |
           _cons |  -.7026203   .2651352    -2.65   0.008    -1.222276   -.1829649
---------------------------------------------------------------------------------
```

Comparing binary choice models

- The real choice is between linear and non-linear models
- Different coefficients across binary choice models do not imply different substantive effects, because they are approximately related:

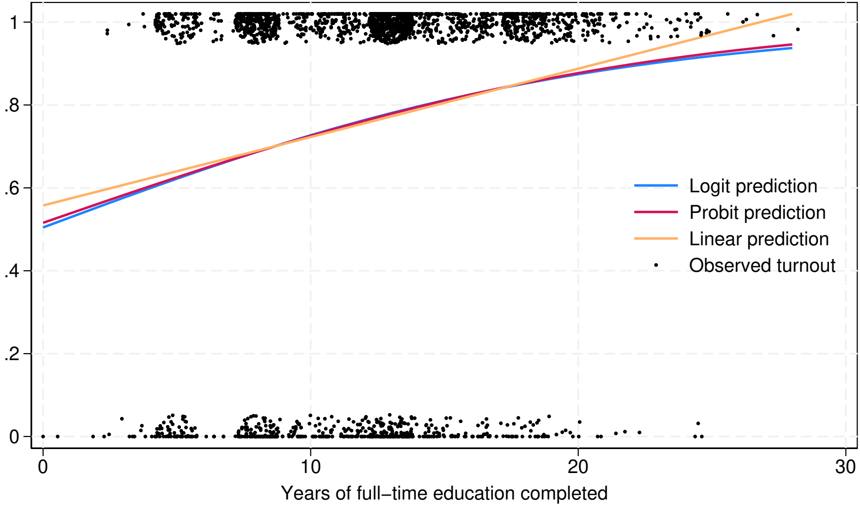
$$\beta_{Logit} \approx 4\beta_{OLS}$$

$$\beta_{Probit} \approx 2.5\beta_{OLS}$$

$$\beta_{Logit} \approx 1.6\beta_{Probit}$$

- Different marginal effects and probabilities from probit and logit arise when extreme values of $x'\beta$ are considered, or the distribution of y is highly skewed
- A set of statistics help identifying the 'best' model:
  - Higher log-likelihood at convergence
  - Higher Pseudo-R^2
  - Higher percentage of correclty classified (`estat classification` after estimation)

Comparing predicted probabilities



Multinomial models

- Now we want to understand how individual characteristics predict voting decisions across different parties (nominal choices)
- The outcome variable presents $k = 1, ..., m$ different categories, each corresponding to a different party
- The binary logic can be extended so as to encompass all choices faced by voting individuals; rather than one, we have m binary variables for each observation y_i

$$y_i = \begin{cases} 1 & \textit{if } y = k \\ 0 & \textit{if } y! = k \end{cases}$$

- Individual's i probability of voting for party k is:

$$p_{i,k} \equiv Pr\left(y_i = k|\boldsymbol{x}_i\right) = F_k\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right)$$

- All probabilities m must sum to one: $\sum_{k=1}^{m} F_k\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right) = 1$

```
fre prtvtdit                          // categorical variable recording the party voted in
       the last general elections (2018); parties are not sorted in any particular order;
      many of them were just voted by a few respondents

generate coalition = .                // create a new variable that aggregates parties
      according to coalition (center-left, center-right, five star, others)
replace coalition = 1 if inlist(prtvtdit,2,7,11,14)     // center-left coalition
replace coalition = 2 if inlist(prtvtdit,3,4,5,8)       // center-right coalition
replace coalition = 3 if inlist(prtvtdit,1)             // five star movement
replace coalition = 4 if coalition== . & prtvtdit < .a  // other minor parties

label variable coalition "Electoral coalition"          // create labels for the new
      variable following the usual procedure
label define coalition_labels 1 "Center-left" 2 "Center-right" 3 "Five star movement" 4 "
      Others"
label values coalition coalition_labels

tab prtvtdit coalition                // check that parties have been sorted correctly


// according to official statistics the five star movement got, as a single party, a
      plurality of votes in 2018; its vote shares were particularly high in Southern and
      Insular Italy
// let's analyze the probability that respondents from Southern and Insular Italy reported
      to have voted for different coalitions

fre region                            // categorical variable recording the Italian NUTS 1
       areas of residence

generate res_south_insular = (region == "ITF" | region == "ITG") if region != ""        //
      fast way to create a dummy variable
label variable res_south_insular "Southern/Insular resident"
```

Multinomial logit model

- $F_k(\cdot)$ can not be a linear function in a multinomial model
- The probability of voting for party k can be modelled using a variation of the logistic function seen before:

$$p_k = Pr(y = j|\boldsymbol{x}) = \frac{e^{\mathbf{x}'\boldsymbol{\beta}_k}}{\sum_{j=1}^{m} e^{\mathbf{x}'\boldsymbol{\beta}_k}}$$

- The multinomial logit model is estimated by ML
- In binary models the choice of either category is naturally opposed to the other: estimation of either probability gives also the other probability
- In multinomial models we have m categories, each with its own probability: it is sufficient to estimate $m-1$ probabilities, but a baseline category must be chosen

```
. mlogit coalition res_south_insular, baseoutcome(1)        // multinomial logit model;
     center-left as baseline

Iteration 0:  Log likelihood =  -1313.532
Iteration 1:  Log likelihood = -1282.4087
Iteration 2:  Log likelihood = -1282.0117
Iteration 3:  Log likelihood = -1282.0117

Multinomial logistic regression                         Number of obs =  1,078
                                                        LR chi2(3)    =  63.04
                                                        Prob > chi2   = 0.0000
Log likelihood = -1282.0117                             Pseudo R2     = 0.0240

------------------------------------------------------------------------------
        coalition | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
------------------+-----------------------------------------------------------
Center_left       |  (base outcome)
------------------+-----------------------------------------------------------
Center_right      |
 res_south_insular |  -.1857657   .1659557    -1.12   0.263    -.5110328    .1395014
            _cons |   .1559128   .0873862     1.78   0.074    -.0153611    .3271867
------------------+-----------------------------------------------------------
Five_star_movement |
 res_south_insular |   .9920254   .1643124     6.04   0.000      .669979    1.314072
            _cons |  -.5232481    .105165    -4.98   0.000    -.7293676   -.3171285
------------------+-----------------------------------------------------------
Others            |
 res_south_insular |  -.1997519   .3688878    -0.54   0.588    -.9227588    .5232549
            _cons |  -2.027326   .1880564   -10.78   0.000    -2.395909   -1.658742
------------------------------------------------------------------------------
```

Multinomial logit model: interpretation

- STATA output includes as many sets of coefficients as the number of included categories of y
- Coefficients of the baseline category of y are set at 0 and used as reference for interpretation
- Coefficients $\widehat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ are not easily interpretable: a one-unit increase in x_k changes y by $p_k\left(\widehat{\beta}_k - \overline{\widehat{\beta}}\right)$ relative to the baseline category; the signs of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{k}}$ do not necessarily provide the directions of relationships
- To report results it is better either to use odds-ratio interpretation (here called relative-risk-ratios), or to compute marginal effects
- Important: changing the baseline category of y affects all coefficients and modifies their interpretation

```
. mlogit coalition res_south_insular, baseoutcome(1) rrr          // multinomial logit model
      with relative-risk-ratios; center-left as baseline

Iteration 0:  Log likelihood =  -1313.532
Iteration 1:  Log likelihood = -1282.4087
Iteration 2:  Log likelihood = -1282.0117
Iteration 3:  Log likelihood = -1282.0117

Multinomial logistic regression                         Number of obs =  1,078
                                                        LR chi2(3)    =  63.04
                                                        Prob > chi2   = 0.0000
Log likelihood = -1282.0117                             Pseudo R2     = 0.0240

------------------------------------------------------------------------------
        coalition |        RRR   Std. err.      z    P>|z|     [95% conf. interval]
------------------+-----------------------------------------------------------
Center_left       |  (base outcome)
------------------+-----------------------------------------------------------
Center_right      |
 res_south_insular |   .8304682   .1378209    -1.12   0.263     .5998757     1.1497
            _cons |   1.168724   .1021304     1.78   0.074     .9847563    1.38706
------------------+-----------------------------------------------------------
Five_star_movement |
 res_south_insular |   2.696691   .4430997     6.04   0.000     1.954196   3.721295
            _cons |   .5925926     .06232    -4.98   0.000     .4822138   .7282372
------------------+-----------------------------------------------------------
Others            |
 res_south_insular |   .8189339   .3020947    -0.54   0.588     .3974211   1.687511
            _cons |   .1316872   .0247646   -10.78   0.000     .0910898   .1903784
------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.
```

```
. mlogit coalition res_south_insular, baseoutcome(3) rrr          // multinomial logit model
      with relative-risk-ratios; five star as baseline

Iteration 0:  Log likelihood =  -1313.532
Iteration 1:  Log likelihood = -1282.4087
Iteration 2:  Log likelihood = -1282.0117
Iteration 3:  Log likelihood = -1282.0117

Multinomial logistic regression                         Number of obs =  1,078
                                                        LR chi2(3)    =  63.04
                                                        Prob > chi2   = 0.0000
Log likelihood = -1282.0117                             Pseudo R2     = 0.0240

------------------------------------------------------------------------------------
         coalition |        RRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------------+----------------------------------------------------------------
Center_left        |
 res_south_insular |   .3708249   .0609311    -6.04   0.000     .2687237    .5117193
             _cons |     1.6875   .1774659     4.98   0.000     1.373179    2.073769
-------------------+----------------------------------------------------------------
Center_right       |
 res_south_insular |   .3079582   .0503223    -7.21   0.000     .2235632    .4242124
             _cons |   1.972222    .201761     6.64   0.000       1.6139    2.410099
-------------------+----------------------------------------------------------------
Five_star_movement |  (base outcome)
-------------------+----------------------------------------------------------------
Others             |
 res_south_insular |   .3036811   .1116781    -3.24   0.001     .1477033    .6243746
             _cons |   .2222222   .0434298    -7.70   0.000     .1515074    .3259426
------------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.
```

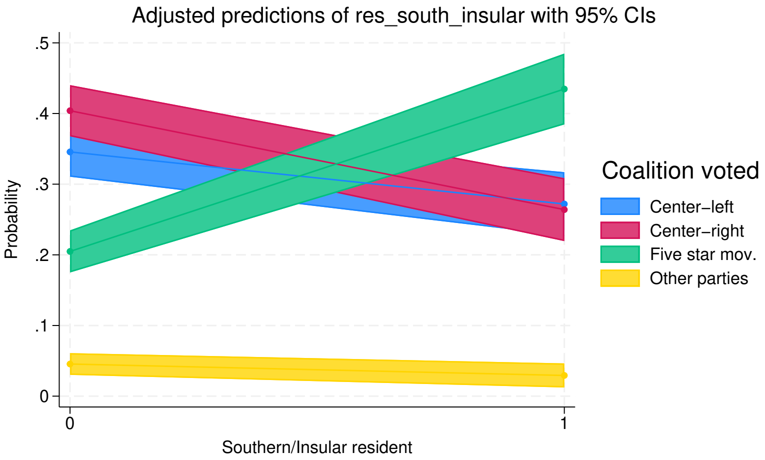
```
. margins, dydx(*)              // marginal effects

Average marginal effects                                   Number of obs = 1,078
Model VCE: OIM

dy/dx wrt: res_south_insular

1._predict: Pr(coalition==Center_left), predict(pr outcome(1))
2._predict: Pr(coalition==Center_right), predict(pr outcome(2))
3._predict: Pr(coalition==Five_star_movement), predict(pr outcome(3))
4._predict: Pr(coalition==Others), predict(pr outcome(4))

------------------------------------------------------------------------------------
                  |            Delta-method
                  |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
------------------+-----------------------------------------------------------------
res_south_insular |
         _predict |
               1  |  -.0624444   .0294574    -2.12   0.034    -.1201798    -.004709
               2  |  -.1319259   .0300395    -4.39   0.000    -.1908022   -.0730495
               3  |   .2097123   .0241428     8.69   0.000     .1623934    .2570312
               4  |   -.015342   .0131253    -1.17   0.242    -.0410672    .0103832
------------------------------------------------------------------------------------
```

Multinomial logit model: interpretation

- Being from Southern or Insular Italy increases the odds of voting for the five star movement by a factor of 2.69 (corresponding to a 169% increase) relative to voting for the center-left coalition
- Or, it decreases the odds of voting for the center-left coalition by a factor of .308 (a 69% decrease) realtive to voting for the five star movement
- Differences in voting across parties other than the five star movement for a resident in Southern or Insular Italy are not statistically significant
- Averaged across all respondents, being from Southern or Insular Italy increase the probability of voting for the five star movement by 21 percentage points, decreases that for the center-left coalition by 6.2 pp, for the center-right by 13.2 pp, for other parties by 1.5 pp

Plot of predicted probabilities

```
margins i.res_south_insular
marginsplot, recastci(rarea) legend(title("Coalition voted") order(1 "Center-
    left" 2 "Center-right" 3 "Five star mov." 4 "Other parties"))
```



Multinomial probit model

- The multinomial logit model imposes the assumption of Independence of Irrelevant Alternatives (IIA): adding one option to choose from must not alter the choice between initial options
- If IIA is unlikely to hold, the multinomial probit model has to be preferred
- The STATA implementation works exactly as the multinomial logit model, except for differently scaled coefficients and for the absence of odds-ratio interpretation (need to compute marginal effects)
- The STATA command is `mprobit`

Ordered choice models

- Suppose we want to understand the determinants of interest in politics, by asking respondents whether they are 'not at all', 'to some extent' or 'very interested'

- The outcome variable y has categories whose order matters: answering 'to some extent' clearly indicates more interest than 'not at all' and less than 'very interested' (ordered choices)

- The outcome variable presents $k = 1, ..., m$ different categories, and it holds that $1 < 2 < ... < k$

- Individual's i probability of reporting interest k is corresponds to intervals of the cumulative distribution function:
  - Logistic: ordered logit model (commmand `ologit`)
  - Standard normal: ordered probit model (coommand `oprobit`)

Ordered choice models

- Coefficients have a difficult interpretation also in these cases, but their signs correctly indicate the direction of effects
- Need to predict probabilities, or predict marginal effects (or provide odds-ratios in the ordered logit case adding the option `, or`)
- An assumption of ordered logit model is the presence of proportional odds: the values attached to different ordered categories are arbitrary, so it is assumed that moving from 0 to 1 is proportional to moving from 1 to 2, etc.
- This assumtpion can be tested with the `brant` command

```
. ologit polintr educ_years i.coalition, or                 // ordinal logit model with odds-
     ratios

Iteration 0:  Log likelihood = -1334.5957
Iteration 1:  Log likelihood = -1274.9161
Iteration 2:  Log likelihood = -1274.3336
Iteration 3:  Log likelihood = -1274.3329
Iteration 4:  Log likelihood = -1274.3329

Ordered logistic regression                             Number of obs =  1,069
                                                        LR chi2(4)    = 120.53
                                                        Prob > chi2   = 0.0000
Log likelihood = -1274.3329                             Pseudo R2     = 0.0452

------------------------------------------------------------------------------------
            polintr | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
--------------------+---------------------------------------------------------------
         educ_years |   1.142903   .0162026     9.42   0.000     1.111584    1.175105
                    |
          coalition |
       Center-right |   .5754308   .0807796    -3.94   0.000     .4370191    .7576799
Five star movement  |   .4689284   .0695468    -5.11   0.000     .3506426    .6271167
             Others |   .6105309   .1843187    -1.63   0.102     .3378552    1.103277
--------------------+---------------------------------------------------------------
              /cut1 |  -.3986842   .2072868                     -.804959    .0075905
              /cut2 |   1.535876   .2108008                     1.122714    1.949038
              /cut3 |   3.831804    .241305                     3.358855    4.304753
------------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation to odds ratios.
```

References

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

———. 2022a. *Microeconometrics Using Stata, Volume I: Cross-Sectional and Panel Regression Methods.* 2nd ed. Stata Press.

———. 2022b. *Microeconometrics Using Stata, Volume II: Nonlinear Methods and Causal Inference Methods.* 2nd ed. Stata Press.

Greene, William H. 2018. *Econometric Analysis.* 8th ed. Pearson.

Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables.* Vol. 7. Sage Publications.

Wooldridge, Jeffrey M. 2012. *Econometric Analysis of Cross Section and Panel Data.* 2nd ed. MIT Press.