Marginal odds ratios

What they are, how to compute them, and why sociologists might want to use them

Ben Jann and Kristian Bernt Karlson

Analytical Sociology: Theory and Empirical Applications
Venice International University, November 14-17, 2022

Outline

Acknowledgements: This presentation is based on work which is part of a project by Kristian Bernt Karlson that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851293).

Background

- Odds ratios form the backbone of much quantitative sociological research.
- Close to a hallmark of the discipline!
- But: Falling out of favor!
  - Magnitude of odds ratios depends on unmeasured covariates orthogonal to the predictor of interest.
  - Noncollapsibility (rescaling bias).
  - Invalidates cross-model and subgroup coefficient comparisons.

Background

- Solutions?
- KHB for cross-model comparisons
- Compare sign not magnitude
- Average marginal effects (AME)
- Linear probability models (LPM)

Background

- AME/LPM might be throwing out the baby with the bathwater, because . . .

  . . . magnitudes depend on the margin

  . . . they focus on absolute probability differences, not relative differences, which is key to much sociological theory and research.

Background

- What we suggest:

  Use **marginal (log) odds ratios**, which . . .

  . . . behave like AME but retain the (relative) odds ratio interpretation!

  - ✔ unaffected by noncollapsibility
  - ✔ an average effect (population-averaged)
  - ✔ comparable across populations/studies

Marginal odds ratios

- Following Zhang (2008) and Daniel et al. (2021) we use potential outcomes notation to define the marginal odds ratio.
- Y_t: Potential outcome that would realize if treatment T was set to level t by manipulation (i.e., without changing anything else).
- We focus on *binary* outcomes only, that is, $Y_t \in \{0, 1\}$ (failure or success).
- Thus:

  $\Pr(Y_t = 1) = E[Y_t]$ is the (marginal) probability that Y_t will be equal to 1 (probability of success).

Marginal odds ratios

- Consider a binary treatment $T \in \{0, 1\}$.
- The marginal odds ratio (MOR) of the alternative treatment ($T = 1$) versus the standard treatment ($T = 0$) is defined as

  $$\mathrm{MOR} = \frac{\upsilon[\Pr(Y_1 = 1)]}{\upsilon[\Pr(Y_0 = 1)]} = \exp\{\ln \upsilon[\Pr(Y_1 = 1)] - \ln \upsilon[\Pr(Y_0 = 1)]\}$$

  where $\upsilon(p) = p/(1-p)$ (odds) and $\ln \upsilon(p) = \ln(p/(1-p))$ (log odds).
- Interpretation of MOR: The ratio of the odds of success if everyone would receive the alternative treatment versus the odds of success if everyone would receive the standard treatment (assuming that there are no general equilibrium effects, i.e., SUTVA holds).

"Marginal" refers to how a predictor affects the "marginal distribution" of an outcome (i.e., not to a marginal change in a predictor). "Unconditional" would be another term but we use "marginal" because the term is established in the literature (Stampf et al. 2010; Karlson, Popham, and Holm 2021).

Adjusting for covariates

- The probability of success may not only depend on T, but also on other factors $\mathbf{X}$.
- Assume that $\mathbf{X}$ has a specific distribution in the population and let $\Pr(Y_t = 1|\mathbf{X} = \mathbf{x}) = E[Y_t|\mathbf{X} = \mathbf{x}]$ be the conditional success probability given $\mathbf{X} = \mathbf{x}$.
- By the law of iterated expectations,

$$\Pr(Y_t = 1) = E_{\mathbf{X}}[\Pr(Y_t = 1|\mathbf{X} = \mathbf{x})]$$

where $E_{\mathbf{X}}$ is the expectation over the distribution of $\mathbf{X}$.

Adjusting for covariates

- The marginal odds ratio, adjusting for **X**, can then be written as

$$\begin{aligned}\text{MOR} &= \frac{\upsilon\{E_{\mathbf{X}}[\Pr(Y_1 = 1|\mathbf{X} = \mathbf{x})]\}}{\upsilon\{E_{\mathbf{X}}[\Pr(Y_0 = 1|\mathbf{X} = \mathbf{x})]\}} \\ &= \exp(\ln \upsilon\{E_{\mathbf{X}}[\Pr(Y_1 = 1|\mathbf{X} = \mathbf{x})]\} - \ln \upsilon\{E_{\mathbf{X}}[\Pr(Y_0 = 1|\mathbf{X} = \mathbf{x})]\})\end{aligned}$$

- We term this the *adjusted* MOR.
- Note:
  - The adjusted MOR is the same as the unadjusted MOR by definition (i.e., same estimand)!
  - However, estimation based on the adjusted MOR formulation can be used to address confounding bias in observational data. It can also be used to increase efficiency in analysis of RCTs.
- The MOR can be defined in a similar way for continuous treatments. For details see our (soon completed) paper.

Relationship to the logistic model

- Consider a simple logistic model

$$\Pr(Y_t = 1) = \text{logit}(\alpha + \delta t) \quad \text{where} \quad \text{logit}(z) = \frac{\exp(z)}{1 + \exp(z)}$$

  which implies

$$\ln \upsilon\{\Pr(Y_t = 1)\} = \alpha + \delta t$$

- Assume T is binary. We then recover the MOR as

$$\text{MOR} = \exp\{(\alpha + \delta) - (\alpha)\} = \exp(\delta)$$

- Meaning: the (exponent of the) slope coefficient in a simple logistic regression estimates the MOR.

Relationship to the logistic model

- If we condition on $\mathbf{X}$, then

$$\ln \upsilon\{\Pr(Y_t = 1|\mathbf{X} = \mathbf{x})\} = \alpha + \delta t + \mathbf{x}\beta$$

- Here $\exp(\delta)$ is the *conditional* odds ratio (i.e., the the odds ratio given a specific value of $\mathbf{X}$).

- The conditional odds ratio (COR) is different from the MOR, which has a more involved form

$$\text{MOR} = \exp(\ln \upsilon\{E_{\mathbf{X}}[\text{logit}(\alpha + \delta + \mathbf{x}\beta)]\} - \ln \upsilon\{E_{\mathbf{X}}[\text{logit}(\alpha + \mathbf{x}\beta)]\})$$

  and which will be different from COR when $\beta \neq \mathbf{0}$.

Relationship to the logistic model

- The difference between MOR and COR is referred to as *noncollapsibility* or *rescaling bias*.

- "Noncollapsibility of the OR derives from the fact that when the expected probability of outcome is modeled as a nonlinear function of the exposure, the marginal effect cannot be expressed as a weighted average of the conditional effects" (Pang et al. 2016).

- MOR will be attenuated compared to COR (what is commonly referred to as rescaling effects).

- But more importantly:

  They correspond to different estimands!
  They are conceptually different.

Why marginal odds ratios?

1. While there exists only one MOR, there are many CORs, as the latter depends on the conditioning set $\mathbf{X}$.
2. Given their "on average" interpretation, MORs are easier to compare across different populations and studies (do not depend on arbitrary conditioning sets).
3. MORs behave like AMEs: They can be compared across different conditioning sets and they are "average" effects implied by a model.

Estimation

- Estimand $\Rightarrow$ Estimation
- There are several approaches how we can estimate the MOR.
  - G-computation (using predictions from a model)
  - Inverse probability weighting
  - Unconditional logistic regression (RIF regression)
- All are discussed in our forthcoming paper (for binary/categorical as well as continuous treatments; including formulas for analytic standard errors).
- Here we focus on G-computation as it closely resembles the formulation of the adjusted MOR above. That is, G-computation obtains the MOR that is *implied* by the chosen logit model. The other methods follow a somewhat different logic.

G-computation

- G-computation estimates the MOR using counterfactual predictions from a logit model (or any other model in principle).
- For example, for a binary treatment, the procedure is as follows.
  1. Regress Y on T and $\mathbf{X}$ using logistic regression.
  2. Use the model estimates to generate two predictions of $\Pr(Y = 1)$ for each observation, one with T set to 0 and one with T set to 1.
  3. Predictions are then averaged across the sample to obtain estimates of the population-averaged success probability by treatment level.
  4. These average predictions can then be plugged into the formula for the MOR:
  $$\ln \widehat{\mathrm{MOR}} = \ln \upsilon(\overline{p}^{T=1}) - \ln \upsilon(\overline{p}^{T=0})$$
- For continuous treatments we have to evaluate level-specific MORs and then average over the treatment distribution. An alternative approach is based on applying fractional logit to counterfactual predictions (this also works for binary/categorical treatments).

Software

- Software implementing the methods is available from GitHub
  - https://github.com/benjann/lnmor
  - https://github.com/benjann/ipwlogit
  - https://github.com/benjann/riflogit

Example

- Application: gender gap in STEM

```
. use stem, clear
(Excerpt from TREE cohort 2)

. describe

Contains data from stem.dta
 Observations:         6,809                  Excerpt from TREE cohort 2
    Variables:             7                  1 Sep 2022 19:28
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
stem            byte    %8.0g                 Is in STEM training
male            byte    %8.0g                 Is male
mathscore       double  %10.0g                Math score
repeat          byte    %8.0g                 Ever repeated a grade
books           byte    %19.0g     books      Number of books at home
wt              double  %10.0g                Sampling weight
psu             int     %8.0g                 Sampling unit
--------------------------------------------------------------------------------
Sorted by:
```

Example

- Probability difference

```
. mean stem [pw=wt], over(male) cluster(psu)
Mean estimation                              Number of obs = 6,809

                   (Std. err. adjusted for 800 clusters in psu)
--------------------------------------------------------------
             |             Robust
             |       Mean   std. err.     [95% conf. interval]
-------------+------------------------------------------------
 c.stem@male |
           0 |    .163234   .0093646      .1448519    .1816161
           1 |   .2748687   .0145161      .2463745    .3033629
--------------------------------------------------------------

. regress stem i.male [pw=wt], cluster(psu) noheader
(sum of wgt is 78,600.1929332293)
                                   (Std. err. adjusted for 800 clusters in psu)
------------------------------------------------------------------------------
             |               Robust
        stem | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      1.male |   .1116347   .0142969     7.81   0.000     .0835708    .1396987
       _cons |    .163234   .0093653    17.43   0.000     .1448506    .1816174
------------------------------------------------------------------------------
```

Example

- Unadjusted (gross) OR

```
. logit stem i.male [pw=wt], or cluster(psu) nolog
Logistic regression                                     Number of obs =  6,809
                                                        Wald chi2(1)  =  67.37
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -40949.278                       Pseudo R2     = 0.0172
                                  (Std. err. adjusted for 800 clusters in psu)
------------------------------------------------------------------------------
             |               Robust
        stem | Odds ratio   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      1.male |   1.943131   .1572663     8.21   0.000     1.658099     2.27716
       _cons |   .1950773   .0133746   -23.84   0.000     .1705485    .2231338
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Example

- Conventional approach "conditional" OR

```
. logit stem i.male mathscore i.repeat books [pw=wt], or cluster(psu) nolog
Logistic regression                                     Number of obs =  6,809
                                                        Wald chi2(4)  = 596.03
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -31905.554                       Pseudo R2     = 0.2343
                                     (Std. err. adjusted for 800 clusters in psu)
```

stem	Odds ratio	Robust std. err.	z	P>|z|	[95% conf.	interval]
1.male	1.959295	.1675426	7.87	0.000	1.65696	2.316794
mathscore	2.606164	.1252437	19.93	0.000	2.371897	2.86357
1.repeat	.6563627	.0965248	-2.86	0.004	.4920011	.8756321
books	1.087051	.0341241	2.66	0.008	1.022185	1.156034
_cons	.1058314	.0166897	-14.24	0.000	.0776926	.1441616

Note: _cons estimates baseline odds.

Example

- G-computation approach (`lnmor` is a post-estimation command, i.e. first estimate the model, then apply `lnmor`)

```
. lnmor i.male, or
Enumerating predictions:
  male[2]..done
Marginal odds ratio                               Number of obs     =      6,809
                                                  Command           =      logit
                                  (Std. err. adjusted for 800 clusters in psu)
------------------------------------------------------------------------------
             |               Robust
        stem | Odds Ratio   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      1.male |   1.677032   .1103015     7.86   0.000     1.473911    1.908145
------------------------------------------------------------------------------
```

Example

- Compare results (SEs in parentheses)

ln(MOR)	Unadjusted	Conditional	Adjusted
1.male	0.664	0.673	0.517
	(0.0809)	(0.0855)	(0.0658)

MOR	Unadjusted	Conditional	Adjusted
1.male	1.943	1.959	1.677
	(0.157)	(0.168)	(0.110)

Example

- Using `at()` to evaluate interactions

```
. probit stem i.male##c.mathscore##c.mathscore##i.repeat##c.books [pw=wt], ///
>     cluster(psu)
```

(*output omitted*)

```
. lnmor i.male, nodots or at(repeat)
Marginal odds ratio                               Number of obs    =     6,809
                                                  Command          =    probit

1: repeat = 0
2: repeat = 1
                                  (Std. err. adjusted for 800 clusters in psu)
```

stem	Odds Ratio	Robust std. err.	t	P>|t|	[95% conf.	interval]
1						
1.male	1.70093	.1258126	7.18	0.000	1.471059	1.966721
2						
1.male	1.525492	.3356916	1.92	0.055	.9904096	2.349659

Example

- Using `at()` to evaluate nonlinear effects

```
. lnmor i.male, nodots or at(mathscore = -2(2)2)

Marginal odds ratio                               Number of obs     =      6,809
                                                  Command           =     probit

1: mathscore = -2
2: mathscore = 0
3: mathscore = 2

                                  (Std. err. adjusted for 800 clusters in psu)
```

stem	Odds Ratio	Robust std. err.	t	P>|t|	[95% conf.	interval]
1						
1.male	1.743237	.7809461	1.24	0.215	.7235215	4.200118
2						
1.male	1.882932	.1949999	6.11	0.000	1.536558	2.307387
3						
1.male	2.041985	.3518247	4.14	0.000	1.456035	2.863737

Example

- Obtain results for several predictors in one call

```
. lnmor i.male mathscore i.repeat books, or
(mathscore has 491 levels; using 82 binned levels)
Enumerating predictions:
  male[2]..mathscore[82]..............................................
  ..........................repeat[2]..books[7].......done
Marginal odds ratio                              Number of obs     =      6,809
                                                 Command           =     probit

                                   (Std. err. adjusted for 800 clusters in psu)
```

stem	Odds Ratio	Robust std. err.	t	P>|t|	[95% conf.	interval]
1.male	1.677096	.1104903	7.85	0.000	1.473649	1.90863
mathscore	2.636666	.1367985	18.69	0.000	2.38136	2.919342
1.repeat	.77176	.0926625	-2.16	0.031	.6097144	.9768729
books	1.061709	.024538	2.59	0.010	1.014619	1.110985

Discussion

- We provide a clear definition of the marginal OR (clarification of estimand).
- We provide flexible software that can estimate the marginal OR for categorical as well as continuous predictors.
- But . . .

  . . . is it worth the hassle? How much do sociologists love odds ratios?

  . . . will it change practice?
- Any other comments/ideas?

References

- Daniel, R., J. Zhang, and D. Farewell. 2021. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal* 63(3): 528–557.
- Karlson, K. B., F. Popham, and A. Holm. 2021. Marginal and Conditional Confounding Using Logits. *Sociological Methods & Research*.
- Pang, M., J. S. Kaufman, and R. W. Platt. 2016. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Statistical Methods in Medical Research* 25(5): 1925–1937.
- Stampf, S., E. Graf, C. Schmoor, and M. Schumacher. 2010. Estimators and confidence intervals for the marginal odds ratio using logistic regression and propensity score stratification. *Statistics in Medicine* 29: 760–769.
- Zhang, Z. 2008. Estimating a Marginal Causal Odds Ratio Subject to Confounding. *Communications in Statistics – Theory and Methods* 38(3): 309–321.