How to assess the fit of multilevel logit models with Stata? A project in progress

2020 Swiss Stata Conference at the University of Bern, November 19th, 2020

"Models should not be true but it is important that they are applicable." (John W. Tukey)

Dr. Wolfgang Langer
Martin-Luther-Universität
Halle-Wittenberg
Institut für Soziologie

Associate Assistant Professor
Université du
Luxembourg

Contents

1. What is the problem ?

Current situation in applied research:

- An increasing number of people use multilevel logistic models for qualitative dependent variables with binary and ordinal outcome
- But users often complain that there are no fit measures for these models
- Neither Stata 16 nor SPSS 26 offer any fit measure for these models
- Let me demonstrate how to generalize the Pseudo R^2s for binary and ordinal logit model for the multilevel analysis

Which solutions does Stata provide?

- Indeed Stata estimates multilevel logit models for binary, ordinal and multinomial outcomes (melogit, meologit, gllamm) but it does not calculate any Pseudo R^2. It provides only the Akaike- (AIC) and Schwarz-Bayesian-Information Criteria (BIC)
- Stata provides a Wald test for the fixed effects and a Likelihood-Ratio-χ^2 test for the random effects of the exogenous variables
- Even special purpose programs like HLM, MlwiN, MPLUS or SuperMix do not calculate any Pseudo R^2

What can we learn from multilevel literature?

- Raudenbush & Bryk (2002), Heck & Thomas (2009) and Rabe-Hesketh & Skrondal (2013) do not mention Peudo R^2s at all
- Snijder & Bosker (2012) propose a variation of McKelvey & Zavoina Pseudo R^2 for random-intercept and intercept-as-outcome logit models. It is not implemented in any program
- Hox (2010) discusses the McFadden, Cox & Snell, Nagelkerke and McKelvey & Zavoina Pseudo R^2. He recommends the last one to assess the model fit

2. Summary of the econometric Monte-Carlo studies for testing Pseudo R^2s

- Econometricians made a lot of Monte-Carlo studies in the 1990s:
  - Hagle & Mitchell 1992
  - Veall & Zimmermann 1992, 1993, 1994
  - Windmeijer 1995
  - DeMaris 2002
- They systematically tested the most common Pseudo-R^2s for binary and ordinal probit / logit models

Which Pseudo R^2s were tested in these studies?

- Likelihood-based measures:
  - Maddala / Cox & Snell Pseudo R^2 (1983 / 1989)
  - Cragg & Uhler / Nagelkerke Pseudo R^2 (1970 / 1992)
- Log-Likelihood-based measures:
  - McFadden Pseudo R^2 (1974)
  - Aldrich & Nelson Pseudo R^2 (1984)
  - Aldrich & Nelson Pseudo R^2 with the Veall & Zimmer-mann correction (1992)
- Basing on the estimated probabilities:
  - Efron / Lave Pseudo R^2 (1970 / 1978)
- Basing on the variance decomposition of the estimated Probits / Logits:
  - McKelvey & Zavoina Pseudo R^2 (1975)

Results of the Monte-Carlo-Studies for binary and ordinal logits or probits

- The McKelvey & Zavoina Pseudo R^2 is the best estimator for the “true R^2” of the OLS regression
- The Aldrich & Nelson Pseudo R^2 with the Veall & Zimmermann correction is the best approximation of the McKelvey & Zavoina Pseudo R^2
- Lave / Efron, Aldrich & Nelson, McFadden and Cragg & Uhler Pseudo R^2 severely underestimate the “true R^2” of the OLS regression
- My personal advice:
  - Use the McKelvey & Zavoina Pseudo R^2 to assess the fit of binary and ordinal logit models

3. The generalization of the McKelvey & Zavoina Pseudo R^2 for the binary and ordinal multilevel logit model

- The multilevel logit model is a systematic extension of the classical binary and ordinal logit model for clustered subsamples (contextual units j)

  - The variance of the estimated logits is decomposed into ▸ Fixed effects, ▸ Random effects and ▸ Level-1 Error variance $\sigma^2(r_{ij})$
  - The variance of level 1 residua $\sigma^2(r_{ij})$ can not be estimated because of its own heteroscedasticity. It is replaced by the variance of the logistic density function ($\pi^2 / 3$) multiplied with the sample size

Let's have a short look at the lucky winner

- McKelvey & Zavoina Pseudo R^2 (M & Z Pseudo R^2)

$$M\ \&\ Z\ Pseudo\ R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i^* - \overline{\hat{y}^*}\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i^* - \overline{\hat{y}^*}\right)^2 + n \times \frac{\pi^2}{3}}$$

Range: $0 \leq$ M & Z-Pseudo $R^2 \leq 1$

Legend:

$\sum_{i=1}^{n}\left(\hat{y}_i^* - \overline{\hat{y}^*}\right)^2$: Sum of squares of the estimated logits (latent variable Y^*)

n : Sample size

$\frac{\pi^2}{3}$: Variance of the logistic density function

Generalization to the 2-level logit model 2

- Prediction of the latent variable Y* (estimated binary or cumulative logit) in two ways
  - 1.Population-Average Prediction with the fixed effects of the exogenous variables (all random effects hold at zero)
    - Stata-command: predict newvar1 if e(sample), xb
  - 2. Unit-Specific Prediction of the fixed and random effects of the exogenous variable
    - Stata-command: predict newvar2 if e(sample), eta
- Therefore, the variation of the estimated logits (Y*) can be calculated in two different ways
  - 1. Only for the fixed effects of the exogenous variables
  - 2. For the fixed and random effects of the exogenous variables

Generalization to the 2-level logit model 3

- Therefore we get two different McKelvey & Zavoina Pseudo R^2s
  - 1. "Population-Average" M & Z Pseudo R^2 (fixed effects)
  - 2. "Unit-Specific" M & Z Pseudo R^2 (fixed & random effects)

- The "Unit-Specific" M & Z Pseudo R^2 uses all estimated fixed and random effects for prediction. Therefore it assesses the fit more realistically as its "Population-Average" counterpart

Let’s have a short look at the lucky loser

- McFadden Pseudo R^2 (1974)

$$McFadden\ Pseudo\ R^2\left(\rho^2\right) = 1 - \left[\frac{\ln L_A}{\ln L_0}\right]$$

Range: $0 \leq$ McFadden Pseudo $R^2 < 1$

but ρ^2 does not reach the maximum of 1.0

Rule of thumb: $0.20 \leq$ McFadden Pseudo $R^2 \leq 0.40$ marks an excellent fit (McFadden 1979: 307)

Legend: $\ln L_A$: Log-Likelihood of the actual model
$\ln L_0$: Log-Likelihood of the zero model

Generalization to the 2-level logit model 4

- Conditions of application
  - Maximum-Likelihood estimation of the fixed and random effects of the exogenous variables
  - Actual and zero model have to use the same sample
  - Choice of the "appropriate zero model" (M_0) depends on our knowlege to which contextual unit the respondent belongs
    - Membership known: Random-Intercept-Only Logit model estimates the proportion of Y* which can be maximally explained by the contextual units (= ANOVA model)
    - Membership unkown: Fixed-Intercept-Only Logit model estimates only the marginal distribution of Y* (= true zero model)

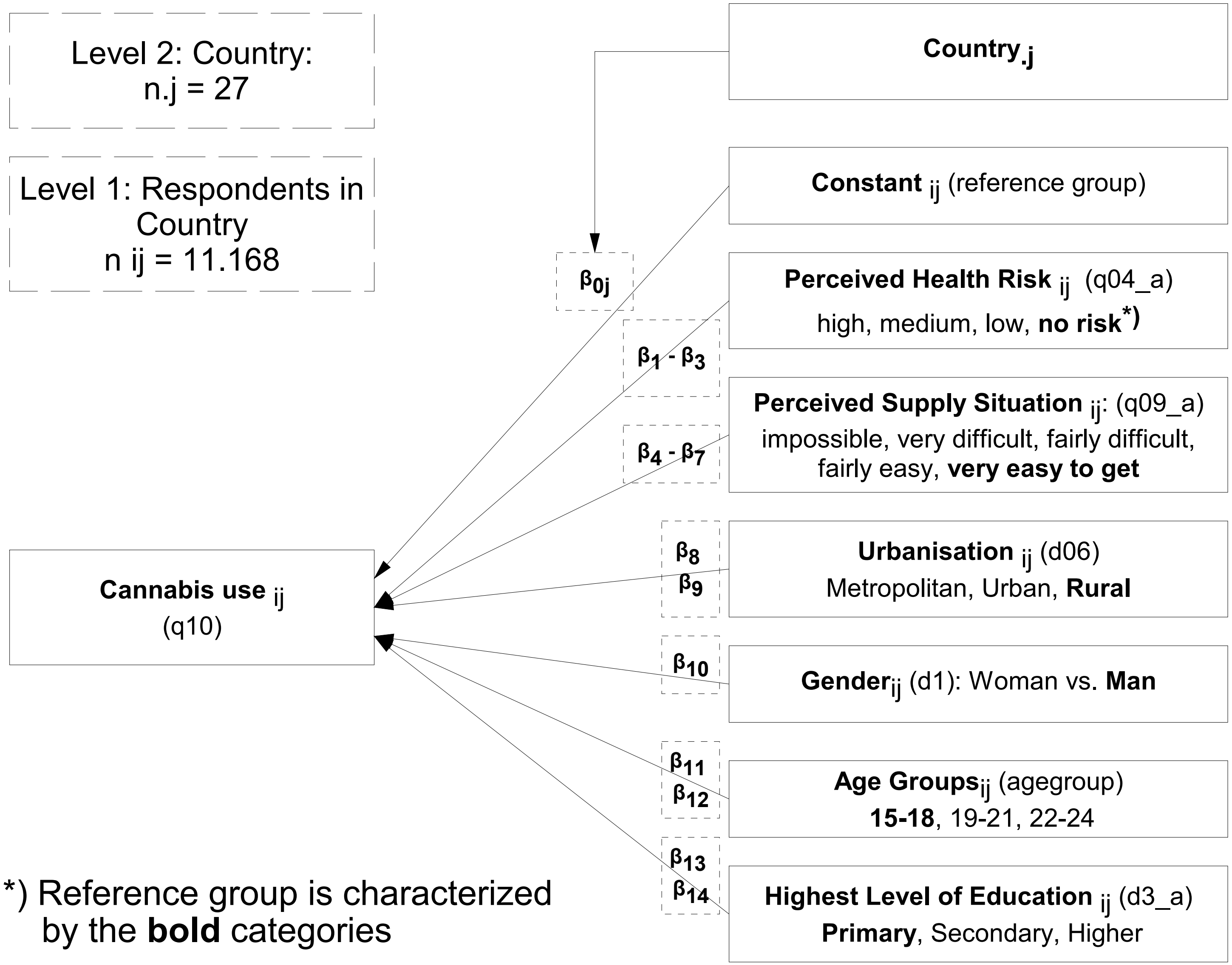
Generalization to the 2-level logit model 5

- Calculation of McFadden Pseudo R^2 is possible in two different ways using the following ones as zero model
  - 1. Random-Intercept-Only Logit-Model (RIOM)
    - It measures the proportional reduction of the log likelihood of the actual model in comparison with the RIOM caused by the fixed effects of the exogenous variables
    - Its Likelihood-Ratio χ^2 test refers to all fixed effects of the exogenous level 1 and level 2 variables
  - 2. Fixed-Intercept-Only Logit-Model (FIOM)
    - It measures the proportional reduction of the log likelihood of the actual model in comparison with the FIOM caused by fixed and random effects of all exogenous variables
    - Its Likelihood-Ratio χ^2 test refers to all fixed and random effects of the exogenous level 1 and level 2 variables

4. Example of application

- Flash Eurobarometer No 330 about youth attitudes on drugs (2011)
  - WebCATI-Survey of n_{ij} = 12.313 respondents (aged 15 -24) in $n_{.j}$ = 27 EU member states (contextual units j)
  - My focus:
    - prevalence of cannabis use by juveniles and young adults (q10): Have you used cannabis by yourself?
      - **1) never**
      - **2) more than 12 months ago**
      - **3) less than 12 months ago**
      - **4) in the last 30 days**
  - Let us have a look at the exogenous variables in the following diagram

Theoretical 2-level-model: RIM



*) Reference group is characterized by the **bold** categories

Stata-Output Version16

- Fixed effects
- Thresholds
- Random effect

```
Mixed-effects ologit regression                 Number of obs     =     11,168
Group variable:         country                 Number of groups  =         27

                                                Obs per group:
                                                              min =        211
                                                              avg =      413.6
                                                              max =        490

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(14)     =    3363.39
Log pseudolikelihood = -7424.2751               Prob > chi2       =     0.0000
                                     (Std. Err. adjusted for 27 clusters in country)
------------------------------------------------------------------------------------
                         |               Robust
                  q10ord |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------
                    q4_a |
               high risk |  -2.656545   .1498799   -17.72   0.000    -2.950304   -2.362785
             medium risk |  -1.668222   .1015971   -16.42   0.000    -1.867348   -1.469095
                low risk |  -.7527713   .0463199   -16.25   0.000    -.8435567   -.6619859
                         |
                    q9_a |
              impossible |  -2.976197   .1910924   -15.57   0.000    -3.350731   -2.601663
          very difficult |  -2.132899   .1642871   -12.98   0.000    -2.454896   -1.810903
        fairly difficult |  -1.527717   .1004747   -15.21   0.000    -1.724644   -1.330791
             fairly easy |  -.6241175   .0843823    -7.40   0.000    -.7895038   -.4587312
                         |
                      d6 |
       metropolitan zone |   .3950294   .1038006     3.81   0.000      .191584    .5984749
other town/urban centre  |   .2082467   .0751678     2.77   0.006     .0609206    .3555729
                         |
                      d1 |
                  female |  -.4777186   .0525238    -9.10   0.000    -.5806634   -.3747738
                         |
                agegroup |
                 19 - 21 |   .4967552   .0607537     8.18   0.000     .3776801    .6158303
                 22 - 24 |   .6850326   .0788467     8.69   0.000     .5304958    .8395694
                         |
                    d3_a |
     secondary education |  -.0122673   .0702398    -0.17   0.861    -.1499348    .1254003
        higher education |  -.0268133   .1227001    -0.22   0.827    -.2673011    .2136745
-------------------------+----------------------------------------------------------
                   /cut1 |  -.3882657   .1392505                     -.6611918   -.1153396
                   /cut2 |   .7153171   .1438259                      .4334236    .9972106
                   /cut3 |   1.902703   .1602572                      1.588605    2.216802
-------------------------+----------------------------------------------------------
country                  |
               var(_cons)|   .2617043   .0850279                      .1384379    .4947284
------------------------------------------------------------------------------------
```

What does Stata offer to assess the fit?

- Akaike (AIC) and Schwarz Bayesian Information Criterion (BIC)
  - Decision rule: Choose the model with the lowest AIC or BIC

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
fiom_w4	11,168	.	-9343.799	3	18693.60	18715.56
riom_w4	11,168	.	-9036.540	4	18081.08	18110.36
rim_w4	11,168	.	-7424.275	18	14884.55	15016.32

Note: BIC uses N = number of observations. See [R] BIC note.

  - Looking at AIC and BIC, the RIM fits best of all bad models
  - But we do not know how well the RIM really fits !

Output of my fit_meologit_2lev.ado 1

- Assessing the fit by the McKelvey & Zavoina-, McFadden-Pseudo R^2s for the fixed & random effects

```
. fit_meologit_2lev
Fit-measures for the MELOGIT/MEOLOGIT Model:

McKelvey&Zavoina-Pseudo R2 (fixed & random effects)=   0.5097

McKelvey&Zavoina-Pseudo R2 (fixed effects only)=   0.4728


Just estimating the Fixed-/ Random-Intercept-Only-Logit Model


McFadden Pseudo R2 (fixed effects only) = 0.1784

McFadden Pseudo R2 (fixed & random effects)  = 0.2054
```

Output of my fit_meologit_2lev.ado 2

- Intra-Class Correlation and corresponding Likelihood-Ratio-χ^2 tests for fixed & random effects

```
ICC of Random-Intercept-Only-Logit Model (Sample M(A))
Intra-Class-Correlation (Level 2) =    0.1431

H0: ICC of Level 2 is zero in the population
LR-chi2 test statistic (  1) =   614.52    Prob > chi2 = 0.0000

LR-chi2 test: H0: all fixed effects are zero in the population

LR-chi2 test statistic ( 14) =  3224.53    Prob > chi2 = 0.0000

LR-chi2 test: H0: all fixed & random effects are zero in the population
LR-chi2 test statistic ( 15) =  3839.05    Prob > chi2 = 0.0000
```

What you get afterwards

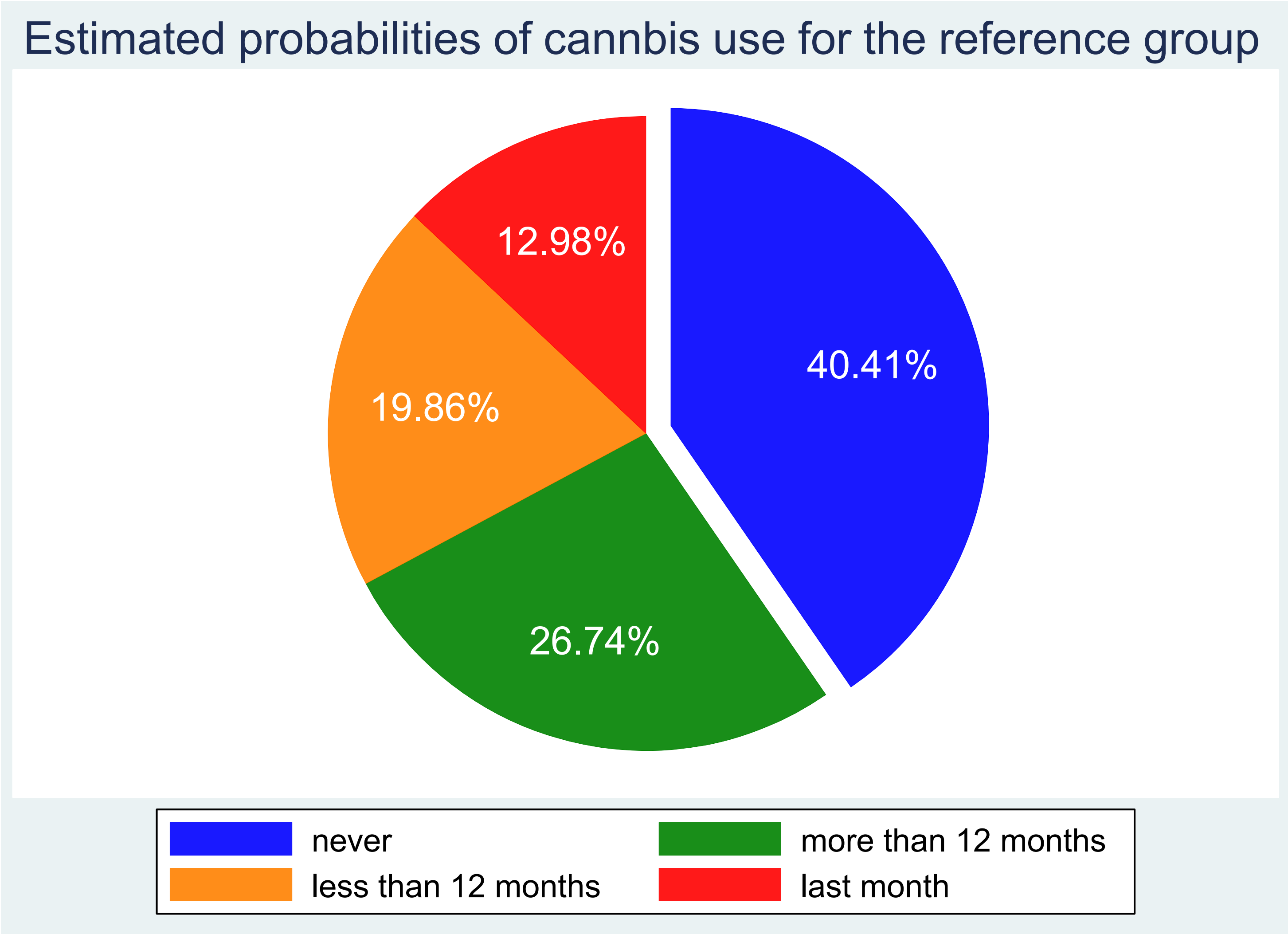
- r-containers of fit_meologit_2lev.ado

```
. return list

scalars:
          r(p_chi2_fr) =  0
              r(df_fr) =  15
            r(chi2_fr) =  3839.047042018567
           r(p_chi2_f) =  0
               r(df_f) =  14
             r(chi2_f) =  3224.530363760103
              r(p_icc) =  0
             r(df_icc) =  1
           r(chi2_icc) =  614.5166782584638
           r(icc_riom) =  .1431007329841504
          r(mcr2_fiom) =  .2054328872219874
          r(mcr2_riom) =  .1784162011068596
              r(mzr2f) =  .4728218379947466
             r(mzr2fr) =  .5097289342586476
```

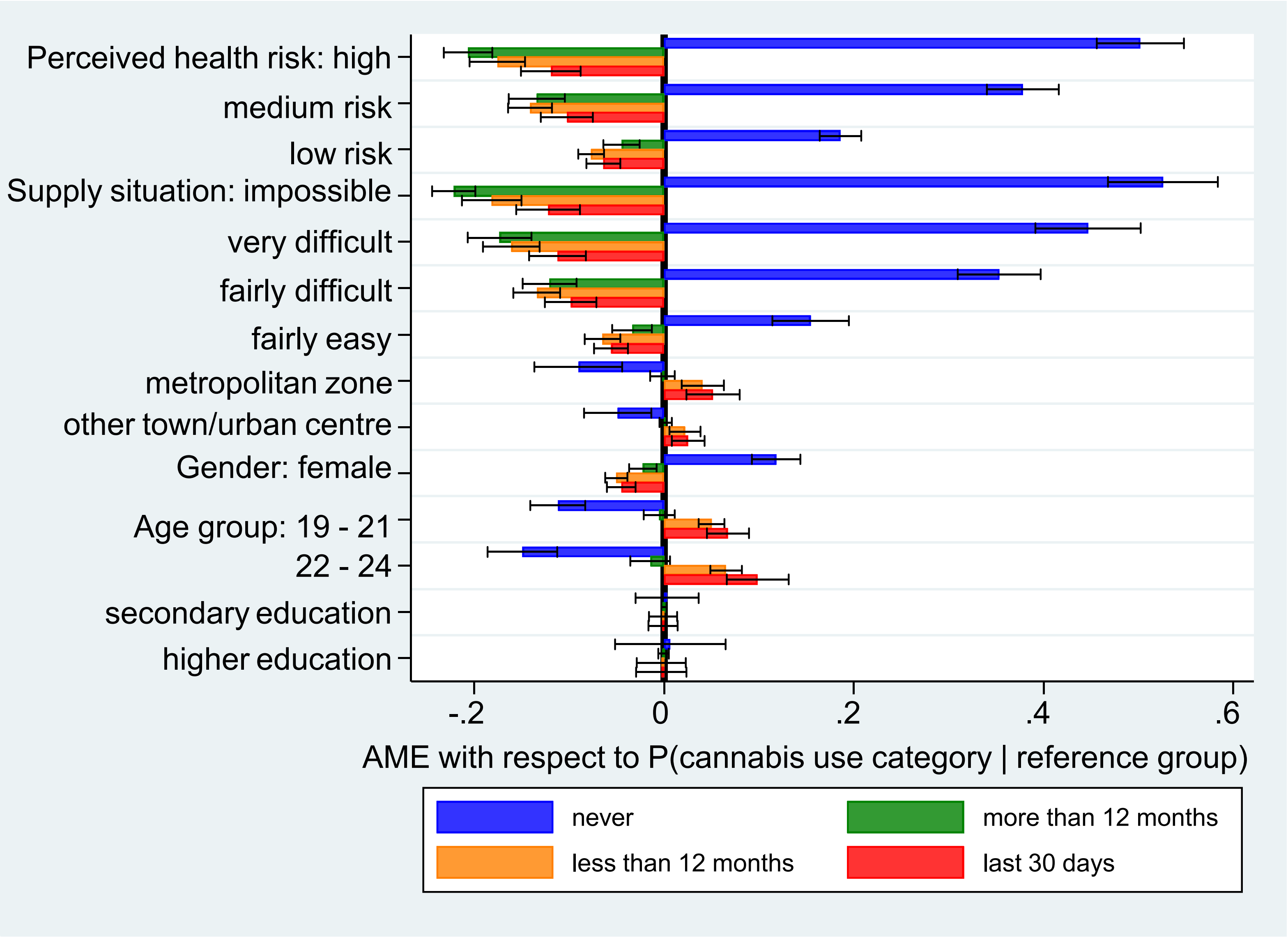
How do the effects look like?

- The baseline



- Reference Group
  - Men
  - Age 15-18
  - Location: rural
  - Education: primary
  - Perceived health risk: no risk
  - Perceived supply situation: very easy to get

Ben Jann's Coefplot for the 4 categories



5. Conclusions 1

- What's known
  - The Monte-Carlo-simulation studies show that the McKelvey & Zavoina Pseudo R^2 is the best fit measure for binary and ordinal logit models
- What's new
  - Generalization of the M & Z-Pseudo R^2 to binary and ordinal multilevel logit models. The prediction of estimated logits bases upon the fixed effects only or upon fixed and random effects of exogenous variables
  - The McFadden-Pseudo R^2 bases upon the fixed effects only or upon fixed and random effects of the exogenous variables using a context-independent zero model

5. Conclusions 2

- What's new
  - Simultaneous Likelihood-Ratio-χ^2 test for the estimated fixed effects using the Random-Intercept-Only Model (RIOM) as the zero model
  - Simultaneous Likelihood-Ratio-χ^2 test for the estimated fixed and random effects using the Fixed-Intercept-Only Model (FIOM) as the zero model
  - Use of probability weights for each level j
  - You get all Pseudo-R2s and tests in r-containers
- That's why
  - I suggest to use my fit_meologit_2lev.ado and fit_meologit_3lev.ado to assess the fit of 2- and 3-level logit models with binary and ordinal outcome

Closing words

- Thank you for your attention
- Do you have some questions?

Contact

- Affiliation
  - Dr.Wolfgang Langer
    University of Halle
    Institute of Sociology
    D-06099 Halle (Saale)
  - Email: wolfgang.langer@soziologie.uni-halle.de

Appendix

Multilevel ordered logit model 1

- Equations of the 2-level-ordered logit model

Level 2: Between-Context Regression

2a) Logistic Intercept-as-Outcome Model:

$$\beta_{0j} = 0 + \gamma_{01} \times Z_{.j} + u_{0j}$$

2b) Logistic Slope-as-Outcome Model:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \times Z_{.j} + u_{1j}$$

Level 1: Within-Context Regression

1) $$\ln\left[\frac{\mathrm{P}(\mathrm{Y} > k)}{\mathrm{P}(\mathrm{Y} \leq k)}\right] = \beta_{0j} + \beta_{1j} \times X_{ij} - \sum_{K=1}^{k-1} \delta_k \{+r_{ij}\}$$

Notation of Raudenbush & Bryk (2002):

γ: fixed-effect estimator
Z: exogenous level 2 variable
β: random-effect estimator
X: exogenous level 1 variable
u_{0j}: residuum random-intercept
u_{1j}: residuum random-slope
r_{ij}: residuum of within-context-logistic regression
δ_k: threshold for category k of Y

Single equation notation: 2a) and 2b) in 1)

$$\ln\left[\frac{\mathrm{P}(\mathrm{Y} > k)}{\mathrm{P}(\mathrm{Y} \leq k)}\right] = \left(0 + \gamma_{01} \times Z_{.j} + u_{0j}\right) + \left(\gamma_{10} \times X_{ij} + \gamma_{11} \times X_{ij} \times Z_{.j} + u_{1j} \times X_{ij}\right) - \sum_{k=1}^{K-1} \delta_k \{+r_{ij}\}$$

Multilevel ordered logit model 2

- Interpretation of the residua of the Between-Context Regression

$$3a)\ u_{0j} = \beta_{0j} - \left[\gamma_{00} + \gamma_{01} \times Z_{.j}\right] = \beta_{0j} - \widehat{\beta_{0j}}$$

$$3b)\ u_{1j} = \beta_{1j} - \left[\gamma_{10} + \gamma_{11} \times Z_{.j}\right] = \beta_{1j} - \widehat{\beta_{1j}}$$

- Assumptions for the residua of the logistic 2-level logit model

Level 1:

1.1) r_{ij} is binomial distributed with an expected value of zero

and a variance $\sigma^2_{r_{ij}} = \widehat{P_{ij}}(Y=1) \times \left(1 - \widehat{P_{ij}}(Y=1)\right)$

1.2) Heteroscedasticity of r_{ij} in all contextual units j

Multilevel ordered logit model 3

- Implication for the level 1 residuum r_{ij}
  - The variance $\sigma^2(r_{ij})$ can not be estimated because of its own heteroscedasticity. It is replaced by the variance of the logistic density function ($\pi^2 / 3$)

- Residua of level 2

2.1) u_{kj} is normal distributed with an expected value of zero and a covariance matrix T of the residua

$$E\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad T = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \qquad \sigma^2_{u_{0j}} = \tau_{00} \qquad \sigma^2_{u_{1j}} = \tau_{11}$$

$$\sigma_{u_{0j},u_{1j}} = \tau_{10} = \tau_{01}$$

2.2) The residua of level 1 and level 2 are not correlated:

$$\sigma_{u_{0j},r_{ij}} = \sigma_{u_{1j},r_{ij}} = 0$$

Alternative in Stata: Information criteria

- Calculation of Akaike- (AIC) and Schwarz Bayesian-Information-Criteria (BIC)

$$AIC = -2 \times \ln L_{M_A} + 2 \times k$$

$$BIC = -2 \times \ln L_{M_A} + \ln N \times k$$

deviance ($-2 \times \ln L_{M_A}$)   complexity of the model (k)

Range: $0 < AIC \leq +\infty$

$0 < BIC \leq +\infty$

Legend:

$\ln L_{M_A}$: Log Likelihood of actual model

k : Number of estimated parameters

N : *Sample size*

References

- Aldrich, J.H. & Nelson, F.D. (1984):
  Linear probability, logit, and probit models. Newbury Park: SAGE
  (Quantitative Applications in the Social Sciences, 45)
- Amemiya, T. (1981):
  Qualitative response models: a survey. *Journal of Economic Literature*, 21,
  pp.1483- 1536
- Begg, C.B. & Gray, R. (1984):
  Calculation of polychotomous logistic regression parameters using individualized
  regression. *Biometrika*, 71, pp.11-18
- Ben-Akiva, M. & S.R.Lerman 1991[4] (1985): Discrete choice analysis. Theory and
  application to travel demand. Cambridge, Mass: MIT-Press
- Cox, D.R. & Snell, E.J. (1989): The analysis of binary data. London: Chapman & Hill
- Cragg, S.G. & Uhler, R. (1970):
  The demand for automobiles. *Canadian Journal of Economics*, 3, pp. 386-406
- DeMaris, A. (2002):
  Explained variances in logistic regression. A Monte Carlo study of proposed
  measures. *Sociological Methods & Research*, 11, 1, pp. 27-74
- Efron, B. (1978):
  Regression and Anova with zero-one data. Measures of residual variation. *Journal
  of American Statistical Association*, 73, pp. 113-121
- Hagle, T.M. & Mitchell II, G.E. (1992):
  Goodness of fit measures for probit and Logit. *American Journal of Political
  Science*, 36, 3, pp. 762-784
- Heck, R.H. & Thomas, S.L. (2009): An Introduction to Multilevel Modeling Techniques.
  New York, N.Y.: Routlege

References 2

- Hensher, D.A. & Johnson, L.W. (1981): Applied discrete choice modelling. London: Croom Helm
- Hensher, D.A., Rose, J.M. & Greene, W.H. (2005): Applied choice analysis. A primer. Cambridge: Cambridge University Press
- Hox, J.J. (2010²): Multilevel Analysis. Techniques and Applications. New York, NY: Routledge
- Huq, N.M. & Cleland, J. (1990): Bangladesh Fertility Surey 1989 (Main Report). Dhaka: National Institute of Population Research and Training
- Long, J.S. (1997): Regression models for categorical and limited dependent variables. Thousand Oaks, Ca : Sage
- Long, J.S. & Freese, J. (2000): Scalar measures of fit for regression models. Bloomington: Indiana University
- Long, J.S. & Freese, J. (2003²): Regression models for categorical dependent variables using Stata. College Station, Tx: Stata Press
- Maddala, G.S. (1983): Limited-dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press
- McFadden, D. (1979): Quantitative methods for analysing travel behaviour of individuals: some recent developments. In: Hensher, D.A. & Stopher, P.R.: (eds): Behavioural travel modelling. London: Croom Helm, pp. 279-318
- McKelvey, R. & Zavoina, W. (1975): A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4, pp. 103-20
- Nagelkerke, N.J.D. (1991): A note on a general definition of the coefficient of determination. *Biometrika*, 78, 3, pp. 691-693

References 3

- Ronning, G. (1991): Mikro-Ökonometrie. Berlin: Springer
- Rabe-Hesketh, S. & Skrondal, A. (2012³): Multilevel and Longitudinal Modeling Using Stata. Volume II: Categorical Responses, Counts, and Survival. Collage Station, Tx: Stata Press
- Raudenbush, S.W. & Bryk, A.S. (2002²): Hierarchical Linear Models. Applications and Data Analysis Methods. Thousand Oaks, CA: Sage
- Snijders, T.A.B. & Bosker, R.J. (2012²): Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, CA: Sage
- Veall, M.R. & Zimmermann, K.F. (1992): Pseudo-R2 in the ordinal probit model. *Journal of Mathematical Sociology*, 16, 4, pp. 333-342
- Veall, M.R. & Zimmermann, K.F. (1994): Evaluating Pseudo-R2's for binary probit models. *Quality & Quantity*, 28, pp. 151- 164
- Windmeijer, F.A.G. (1995): Goodness-of-fit measures in binary choice models. *Econometric Reviews*, 14, 1, pp. 101-116
- Zimmermann, K.F. (1993): Goodness of fit in qualitative choice models: review and evaluation. In: Schneeweiß, H. & Zimmermann, K. (eds): Studies in applied econometrics. Heidelberg: Physika, pp. 25-74