How to assess the fit of multilevel logit models with Stata?

Meeting of the German Stata User Group at the Humboldt University Berlin, June 23rd, 2017

“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
1. What is the problem?
2. Summary of the econometric Monte-Carlo studies for Pseudo R²s
3. The generalization of the McKelvey & Zavoina Pseudo R² for the binary and ordinal multilevel logit model
4. An application of the generalized M&Z Pseudo-and McFadden Pseudo R² in a drug consumption study of juveniles and young adults
5. Conclusions
1. What is the problem?

Current situation in applied research:

- An increasing number of people uses 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 14 / 15 nor SPSS 24 offer any fit measure for these models
- Let me demonstrate how to generalize the Pseudo $R^2$'s 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 information criteria AIC and BIC (estat ic).

- Stata provides a Wald-test for the fixed-effects and a Likelihood-Ratio-$\chi^2$ 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?


- 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 $R^2$s

- Econometricians made a lot of Monte-Carlo studies in the early 90s:
  - Hagle & Mitchell 1992
  - Windmeijer 1995
  - DeMaris 2002

- They tested systematically the most common Pseudo-$R^2$s for binary and ordinal probit / logit models
Which Pseudo $R^2$s were tested in these studies?

- **Likelihood-based measures:**

- **Log-Likelihood-based measures:**
  - McFadden Pseudo $R^2$ (1974)
  - Aldrich & Nelson Pseudo $R^2$ with the Veall & Zimmermann correction (1992)

- **Basing on the estimated probabilities:**

- **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 McElvey & 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 $\sigma^2(r_{ij})$

- Because of its own heteroscedasticity the variance of level 1 residua $\sigma^2(r_{ij})$ can not be estimated. It is replaced by the variance of the logistic density function $(\pi^2 / 3)$
Let’s have a short look at the lucky winner

● McKelvey & Zavoina Pseudo R² (M & Z Pseudo R²)

\[ M \text{ & } Z \text{ Pseudo } R^2 = \frac{Var(\hat{y}^*)}{Var(\hat{y}^*) + Var(\varepsilon)} = \frac{\sum_{i=1}^{n} (\hat{y}_i^* - \bar{y}^*)^2}{\sum_{i=1}^{n} (\hat{y}_i^* - \bar{y}^*)^2 + n \pi^2 / 3} \]

Range: \( 0 \leq M \& Z\text{-Pseudo } R^2 \leq 1 \)

Legend:

- \( Var(\hat{y}^*) \): Variance of the estimated logits (latent variable \( Y^* \))
- \( \hat{y}_i^* \): Estimated logit of case \( i \)
- \( \bar{y}^* \): Expected value of the estimated logits
- \( \pi^2 / 3 \): Variance of the logistic density function
Generalization to the 2-level logit model

- Prediction of the latent variable $Y^*$ (estimated binary or cumulative logit) in two ways
  - 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`
  - Unit-Specific Prediction of the fixed and random effects of the exogenous variable
    - Stata-command: `predict newvar2 if e(sample), eta`

- Therefore, the variance of the estimated logits ($Y^*$) can be calculated in two different ways
  - Only for the fixed effects of the exogenous variables
  - For the fixed and random effects of the exogenous variables
Therefore we get two different McKelvey & Zavoina Pseudo $R^2$'s

- "Population-Average" M & Z Pseudo $R^2$ (fixed effects)
- "Unit-Specific" M & Z Pseudo $R^2$ (fixed- & random effects)

For the "Unit-Specific" M & Z Pseudo $R^2$ uses all estimated fixed and random effects for prediction, 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)

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

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

but $\rho^2$ 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: \( \log L_A \): Log-Likelihood of the actual model
\( \log L_0 \): Log-Likelihood of the zero model
Generalization to the 2-level logit model

Conditions of application

- Maximum-Likelihood estimation of the fixed and random effects of the exogenous variables
- Actual and zero model has to use the same sample
- Choice of the “appropriate zero model” (M₀) depends on our knowledge to which context the respondent belongs
  - **Membership known**: Random-Intercept-Only Logit model estimates the proportion of Y* which can be maximally explained by the context (= ANOVA model)
  - **Membership unknown**: Fixed-Intercept-Only Logit model estimates only the marginal distribution of Y* (= true zero model)
Calculation of McFadden Pseudo $R^2$ is possible in two different ways using the following as a zero model:

- **Random-Intercept-Only Logit-Model (RIOM)**
  - It measures the proportional reduction of the log likelihood of the actual model caused by the fixed effects of the exogenous variables in comparison to the RIOM.
  - Its Likelihood-Ratio $\chi^2$ test refers to all fixed effects of the exogenous level 1 and level 2 variables.

- **Fixed-Intercept-Only Logit-Model (FIOM)**
  - It measures the proportional reduction of the log likelihood of the actual model caused by fixed and random effects of all exogenous variables in comparison to the FIOM.
  - Its Likelihood-Ratio $\chi^2$ 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

Level 2: Country: n.j = 27

Level 1: Respondents in Country n ij = 11,168

Cannabis use ij (q10)

β0j

β1 - β3

β4 - β7

β8

β9

β10

β11

β12

β13

β14

Country.j

Constant ij (reference group)

Perceived Health Risk ij (q04_a)
high, medium, low, no risk

Perceived Supply Situation ij: (q09_a)
impossible, very difficult, fairly difficult,
fairly easy, very easy to get

Urbanisation ij (d06)
Metropolitan, Urban, Rural

Gender ij (d1): Woman vs. Man

Age Groups ij (agegroup)
15-18, 19-21, 22-24

Highest Level of Education ij (d3_a)
Primary, Secondary, Higher
**Stata-Output**

**Version 14**

---

**Fixed effects**

| Fixed effect                  | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-------------------------------|---------|-----------|-------|-------|---------------------|
| q10ord                        |         |           |       |       |                     |
| high risk                     | -2.670499 | .1092326  | -24.45| 0.000 | -2.884591 -2.456407 |
| medium risk                   | -1.696693 | .0730464  | -23.23| 0.000 | -1.839861 -1.553525 |
| low risk                      | -.7425748 | .0611709  | -12.14| 0.000 | -.8624676 -.622682  |
| q9_a                          |         |           |       |       |                     |
| impossible                    | -3.006983 | .1899514  | -15.83| 0.000 | -3.379281 -2.634685 |
| very difficult                | -2.191986 | .1207629  | -18.15| 0.000 | -2.428677 -1.955295 |
| fairly difficult              | -1.555672 | .0870857  | -17.86| 0.000 | -1.726357 -1.384987 |
| fairly easy                   | -.6291072 | .0553719  | -11.36| 0.000 | -.7376341 -.5205803 |
| d6                            |         |           |       |       |                     |
| metropolitan zone             | .3536598 | .0713306  | 4.96  | 0.000 | .2138545 .4934652  |
| other town/urban centre       | .196061  | .0606935  | 3.23  | 0.001 | .0771039 .315018   |
| d1                            |         |           |       |       |                     |
| female                        | -.4654088 | .0504709  | -9.22 | 0.000 | -.5643298 -.3664877 |
| agegroup                      |         |           |       |       |                     |
| 19 - 21                       | .4924681 | .073827   | 6.67  | 0.000 | .3477699 .6371663  |
| 22 - 24                       | .6847313 | .0797637  | 8.58  | 0.000 | .5283974 .8410652  |
| d3_a                          |         |           |       |       |                     |
| secondary education           | -.0345302 | .0753855  | -0.46 | 0.647 | -.1822832 .1132227 |
| higher education              | -.0415283 | .099673   | -0.42 | 0.677 | -.2368837 .1538271 |

**Thresholds**

| Threshold                  | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|----------------------------|---------|-----------|-------|-------|---------------------|
| /cut1                      | -4.269461 | .1329725  | -3.21 | 0.001 | -.6875674 -.1663248 |
| /cut2                      | .6715688 | .133064   | 5.05  | 0.000 | .4107681 .9323695  |
| /cut3                      | 1.857033 | .1357061  | 13.68 | 0.000 | 1.591053 2.123012   |

**Random effect**

| Random effect          | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|------------------------|---------|-----------|-------|-------|---------------------|
| var(_cons)             | .2623196 | .0849424  | .3039| 0.760 | .0998878 .4247501  |

**LR test vs. ologit model:**  
chibar2(01) = 222.09  
Prob >= chibar2 = 0.0000
What does Stata offer to assess the fit?

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

Looking at AIC and BIC, the rim fits best of all bad models. But we do not know how well the rim really fits!
Assessing the fit by the McKelvey & Zavoina-Pseudo R^2's and the Intra-Class-Correlation

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

McKelvey&Zavoina-Pseudo-R2 (fixed&random effects) = 0.5137
McKelvey&Zavoina-Pseudo-R2 (fixed effects only) = 0.4774

Just estimating the Random-/Fixed Intercept Only Logit-Model
Intra-Class-Correlation (Level 2) = 0.1507
```
McFadden Pseudo R²'s and corresponding Likelihood-Ratio-χ² tests

McFadden Pseudo-R2 (M_A vs. Random-Intercept-Only-Logit Model) = 0.1796

McFadden Pseudo-R2 (M_A vs. Fixed-Intercept-Only-Logit Model) = 0.2054

Likelihood-Ratio-chi2-Test (H0: All fixed effects = 0)

Likelihood-ratio test
(Assumption: riom nested in ma) LR chi2(14) = 3245.04
Prob > chi2 = 0.0000

Likelihood-Ratio-chi2-Test (H0: All fixed & random effects = 0)

Likelihood-ratio test
(Assumption: fiom nested in ma) LR chi2(15) = 3832.18
Prob > chi2 = 0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on the boundary of the parameter space. If this is not true, then the reported test is conservative.
How do the effects look like?

The baseline

Estimated probabilities of cannabis use for the reference group

- 39.49% (never)
- 26.7% (more than 12 months)
- 20.31% (less than 12 months)
- 13.51% (last 30 days)
The joint marginsplot for the 4 categories

Conditional Marginal Effects with 95% CIs

Effects with Respect to

- never
- more than 12 months
- less than 12 months
- last 30 days
5. Conclusions

- **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

- **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

● New
  ▶ Simultaneous Likelihood-Ratio-$\chi^2$ test for the estimated fixed effects using the random-intercept-only (RIOM) as the zero model
  ▶ Simultaneous Likelihood-Ratio-$\chi^2$ test for the estimated fixed and random effects using the fixed-intercept-only (FIOM) as the zero model

● 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
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:
\[ \beta_{ij} = \gamma_{10} + \gamma_{11} \times Z_{.j} + u_{1j} \]

Level 1: Within-Context Regression

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

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

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

Interpretation of the residua of the Between-Context-Regression

1. $u_{0j} = \beta_{0j} - \left[ \gamma_{00} + \gamma_{01} \times Z_{.j} \right] = \beta_{0j} - \hat{\beta}_{0j}$
2. $u_{1j} = \beta_{1j} - \left[ \gamma_{10} + \gamma_{11} \times Z_{.j} \right] = \beta_{1j} - \hat{\beta}_{1j}$

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 $\sigma_{r_{ij}}^2 = \hat{P}_{ij}(Y = 1) \times (1 - \hat{P}_{ij}(Y = 1))$

1.2) Heteroscedasticity of $r_{ij}$ in all contextual units j
Multilevel ordered logit model

- Implication for the level 1 residuum $r_{ij}$
  - Because of its own heteroscedasticity the variance $\sigma^2(r_{ij})$ can not be estimated. 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

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

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

$$
\sigma_{u_{0j},r_{ij}} = \sigma_{u_{1j},r_{ij}} = 0
$$
Calculation of Akaike- (AIC) and Schwarz Bayesian-Information-Criteria (BIC)

\[ AIC = -2 \times \log L_{MA} + 2 \times k \]

\[ BIC = -2 \times \log L_{MA} + \log N \times k \]

Legend:

log: \hspace{1cm} \textit{Logarithmus naturalis}

k: \hspace{1cm} \textit{Number of estimated parameters}

N: \hspace{1cm} \textit{Sample size}

Range:

\[ 0 < AIC \leq +\infty \]

\[ 0 < BIC \leq +\infty \]
References


References 2

  Applied discrete choice modelling. London: Croom Helm
  Regression models for categorical and limited dependent variables. Thousand Oaks, Ca: Sage
  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
  Limited-dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press
– McFadden, D. (1979):
  A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology, 4, pp. 103-20
  A note on a general definition of the coefficient of determination. Biometrika, 78, 3, pp.691-693
References 3