Advanced Quantitative Methods in Political Science: How to write a publishable Paper & Multinomial Choice Models

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova

University of Mannheim
School of Social Sciences

Week 9 - 27 April 2022

Intro

What should you take home from this class today?

- You (finally) get my semi-random thoughts on how to write publishable papers.
- We expand our toolbox, again, and learn how to model nominal outcomes.
- We will derive the so-called *multinominal logit* model (MNL) in two different ways.
- We also discuss the so-called *IIA assumption* for MNL (and CL) models if time permits.

Quiz

Quiz

We can extend the ordered probit model to allow for heteroskedasticity. The *heteroskedastic ordered probit* log-likelihood is just:

$$\begin{aligned} lnL(\beta, \tau, \gamma) &= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ji} ln(Pr(Y_{ji} = 1)) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ji} ln(\Phi(\frac{\tau_j - X_i\beta}{e^{Z_i\gamma}}) - \Phi(\frac{\tau_{j-1} - X_i\beta}{e^{Z_i\gamma}})) \end{aligned}$$

Which of the following statements is false?

1. If all elements of γ equal 0, then the model is identified just as in the standard ordered probit model.
2. The model is not identified if a constant is included (i.e. if X_i or Z_i includes a column of 1s).
3. The variance of the error term is a function of a set of predictors in X and Z because of the non-linear probit link.
4. If $J = 1$, then we get a heteroskedastic probit model.

Leftovers:
How to write a publishable Paper

6 Ways to Write a Publishable Paper

1. Go read the journals you wanna get published in!
    - How long are they? How many figures and tables?
    - Do they consist of similar sections or subsections?
2. Be innovative but not too innovative
    - Don't use new theory with new data and new methods
    - Reviewers might not see innovations if everything changes
    - Introduce only one innovation at a time
3. You need to have a point.
    - What exactly is your contribution?
    - *Whose mind are you going to change about what?* (Gary King)
4. Prose must be an "easy read"
    - Use clear and concise language
    - Well-structured article
5. Have a sexy title and a concise abstract
6. Tell a good story
    - Avoid subplots
    - Make that clear in the introduction

Writing Style

- Clarity trumps!
    - Get rid of words that do not change the content of the sentence
    - Avoid pseudo-scientific jargon, acronyms and abbreviations
    - Use active voice
    - Use short sentences (10 words on average, 25 max)
- Use sections and subsections to organize your text
    - Descriptive headings (summarizing key points)
    - Hypothetical table of content should already convey your argument
    - Including section breaks help readers to skip stuff without getting lost
    - Short summaries and appropriate transitions
    - Imagine a reader falling asleep for a few minutes while turning pages!
- Get you message across! Allocate the space accordingly
    - Should not reflect the time you spent accomplishing it
    - I worked once two weeks for a sentence that ended up in a footnote

How to organize your paper

- Common Structure. Segmented in …
    - Title
    - Abstract
    - Introduction
    - Literature Review
    - Theory
    - Research Design
    - Analysis
    - Conclusion
- General Rule: Follow the “Kosher Principle” (Jim Stimson)
    - Content of segments ought to never touch one another
    - Each segment possesses different purpose
    - Thus, do not discuss literature in same section as your analysis,…

Purpose of your Title

1. Convey your point in one phrase
2. Draw attention to your paper
3. Motivate readers to glance over abstract

Purpose of your Abstract

1. Drive home to readers your main point (take-home message)
2. Descriptive overview of project

- One-paragraph, typically $<$ 150 words
- Include information about ...

  ...why is it important
  ...for whom
  ...the results

Purpose of your Introduction

1. Focuses Readers attention by providing a context for what you are about to present
2. Motivates your readers to keep on reading
3. Sell your work

- There is only one first sentence (paragraph). Get it right!
  - Start with explicit question you're going to address
  - Or introduce a puzzle
- Then build a (persuasive) case that this is an interesting and important question, to which the reader needs to find out your answer.
- Clarify the "stakes" of your analysis
  - What is your contribution and why is it important
  - Why should scholars care how you answer it?
  - What are the broader implications of answering it one way or another?
- Readers should be able to come away from the introduction with a strong sense of what you're trying to find out and why it matters

Purpose of your Literature Review

1. Does <u>not</u> *review* the literature
2. Establishes and structures the context of your study
3. It is your chance to frame the issues at stake. Do not let others frame those issues for you!

- Identify a structure in the literature (that fits your goal)
- Summarize different lines of argument, strengths and weaknesses
- Identify a relevant gap in the literature (not every gap is relevant!)
- Relates the gap in the literature to the argument developed in the next section. (No theory discussion here)
- Know your contribution before you write the literature review. Otherwise you will waste your time

Purpose of your Theory Section

1. Presentation of your theoretical contribution

- Intro paragraph summarizing the theoretical argument works well
- Introduce (if not done previously) key concepts
- State your assumptions clearly
- Make sure that your hypotheses follow from your assumption
- Develop your argumentation step-by-step. Start simple, get more complicated.
- Summarize your argument at the end

Purpose of Research Design Section

1. Provides explicit connection between implications of theory and observable facts

- Remember the reader about the purpose of the analysis
- Case Selection
- Time-frame, Period Selection
- Justification of Operationalization
- Description of Data Collection and Coding (if relevant)
- Descriptive Statistics (if necessary)
- Justification of Method and Model Specification. Write down your model (even formalized if not standard)

Purpose of Analysis Section

1. Explain what you did and what you have found

- Need to bring readers along! Don't lose them.
- Description of Results
- Substantive Interpretation using quantities of interest
- Use Graphs (and tables) to make *your* point!
- Do auxiliary analysis: trace observable implications of competing theories
- Do robustness checks: Are results robust to different operationalizations, specifications or methods?

For the final draft paper for this course I would like you to include the plain estimation results of raw coefficients *only* in an Appendix to your paper. Try to write-up an interesting analysis section without it!

Purpose of Conclusion

1. Summarize findings and results effectively
    - Come back to frame from introduction
    - Describe the problem
    - Answer your research question. What has not been answered?
2. Address relevance of your results
3. Discuss implications of your results
    - Point to emerging issues or repeated patterns
    - How do your findings change the way wethink about these issues?
    - Shortcomings of your studies
    - What was your take-home message again?

Final Paper

Expectations - Final Paper (No Replication Report!)

1. The paper should apply or develop an appropriate statistical model to an important substantive problem. This implies the following:
    - The draft paper must include all analysis, tables, figures, and description of the results.
    - A good write-up of the draft paper should read like the third quarter of a journal article.
    - Also have an introduction that makes clear why your topic is important, and outlines what the contribution of this paper will be.
    - The rest of the draft may be in detailed outline form, although it would be better to have it fully written.
    - Think of it as a draft research paper that you are going to submit over the summer to a journal for publication.
2. Provide all necessary information to replicate your analysis!
    - The replication material must include your dataset and computer code to be able to reproduce all tables and figures that make it in the paper.
    - Comment your computer code heavily to explain what you are doing.
    - Your code must be neatly formatted and run cleanly. To that end, please avoid writing computer-specific lines into your code that will prevent it from running on other machines. I will award partial credit if necessary.

In case you start with a replication …

- Do not replicate the entire article.
- Replicate something important. Why is it important? Not because of the authors say so but because *you* say so! This might be something different (e.g. a different dependent variable).
- You have to make a case that this is important. How do you know? We are writing for an audience. You have to convince others that this is important.
- Even if authors say that the paper is about X you can say we should think about C because it is a more interesting question.
- You need to figure out how to cast an article (big picture) and, equally important, do all the little details of squaring or elegantly parameterizing terms to come up with the likelihood. Don't loose sight of either side!
- Write down your model!
- Don't trust that the model assumptions are true. Test them!

Final Paper Deadlines

- Student presentation of your projects: 1 June 2022.
- The final draft paper together with all replication material are due on June 8^{th}, 2022.
- Please submit all files electronically by 10am that day. Late submissions will not be accepted.

Models for Nominal Outcome Variables

Nominal Response Variables

Categories of a nominal variable cannot be ordered. There are many social outcomes that are nominal, e.g.

- Vote-choice (e.g. CDU/CSU, SPD, Greens, FDP, ...)
- MP's vote in favor or against a proposal or abstains
- A typology (of events or actions)

If the dependent variable is ordered, a model for nominal outcomes can still be used (there is some efficiency loss, though). Using ordinal logit/probit to model nominal outcomes yields biased estimates, though.

The Multinomial Logit Model as a Probability Model

- Let y be the dependent variable with J nominal outcomes $m = \{1, \ldots, J\}$ (not necessarily ordered but mutually exclusive). Those different outcomes are aka *choice-set*.
- Let $Pr(y_i = j) = Pr(y_{ij} = 1) = \pi_{ij}$ be the probability for observation i of choosing outcome j such that $\sum_{j=1}^{J} \pi_{ij} = 1$.

The Multinomial Logit Model as a Probability Model

One can construct a probability model as follows:

1. Assume that $Pr(y_i = j|X_i)$, i.e., be the probability for i of choosing outcome j given X, is a linear combination of $X_i\beta_j$, whereby β_j is a choice-specific vector (including a constant) of the effects of each independent variable on observing outcome j. Note that in contrast to binomial or ordered logit, β differs for each outcome (e.g., the effect of education on the probability to vote FDP is different than the effect of education on the probability to vote SPD).
2. Probabilities need to be non-negative. Thus, we need an appropriate link-function to get from $X_i\beta_j$ to probabilities. Hence, we take $exp(X_i\beta_j)$.
3. Probabilities need to sum to 1. Therefore, we need to normalize by dividing by $\sum_{m=1}^{J} exp(X_i\beta_m)$. Thus,

$$\pi_{ij} = Pr(y_i = j|X_i) = \frac{exp(X_i\beta_j)}{\sum_{m=1}^{J} exp(X_i\beta_m)}$$

Identification of the Multinomial Logit Model

While now $\sum_{j=1}^{J} Pr(y_i = j|X_i) = \sum_{j=1}^{J} \pi_{ij} = 1$ (Do you see that?) the probabilities are *not identified* since more than one set of parameters generates the same probabilities of the observed outcomes.

$$\begin{aligned}
Pr(y_i = j|X_i) &= \frac{exp(X_i\beta_j)}{\sum_{m=1}^{J} exp(X_i\beta_m)} \\
&= \frac{exp(X_i\beta_j)}{\sum_{m=1}^{J} exp(X_i\beta_m)} \cdot \frac{exp(X_i\tau)}{exp(X_i\tau)} \\
&= \frac{exp(X_i\beta_j + X_i\tau)}{\sum_{m=1}^{J} exp(X_i\beta_m + X_i\tau)} \\
&= \frac{exp(X_i(\beta_j + \tau))}{\sum_{m=1}^{J} exp(X_i(\beta_m + \tau))}
\end{aligned}$$

Thus, β_j can be replaced with $\beta_j + \tau$ without changing the probabilities and the model predictions. Hence the model is non-identified. We need to impose constraints to identify it.

Identification of the Multinomial Model

- Typically, the model gets identified by constraining one of the β's to zero.
- The choice is arbitrary. Lets assume $\beta_1 = 0$ as baseline. (dimension?)
- Based on this assumption the model is identified. If we add a nonzero τ to β_1 we do not get the same probabilities.
- Adding this constraint yields the following probability equation
$$Pr(y_i = j|X_i) = \frac{exp(X_i\beta_j)}{\sum_{m=1}^{J} exp(X_i\beta_m)}, \text{ whereby } \beta_1 = 0$$
- Commonly, the probability equation is also written as follows

$$\begin{aligned} Pr(y_i = 1|X_i) &= \frac{1}{1 + \sum_{m=2}^{J} exp(X_i\beta_m)} \\ Pr(y_i = j|X_i) &= \frac{exp(X_i\beta_j)}{1 + \sum_{m=2}^{J} exp(X_i\beta_m)}; m > 1 \end{aligned}$$

Deriving the Likelihood Function

Given that each decision-maker chooses one and only one alternative we use the indicator y_{ji} (i.e., equal to 1 when $y_i = j$, and 0 otherwise) to accomplish this.

Thus, the log-likelihood contribution L_i of decision-maker i is

$$\begin{aligned} lnL_i &= ln(\prod_{j=1}^{J} Pr(y_i = j)^{y_{ij}}) \\ &= \sum_{j=1}^{J} y_{ij} ln(Pr(y_i = j)) \end{aligned}$$

Then summing-up all N individual contributions assuming independent realizations gives us the log-likelihood of the multinomial logit model with $J-1$ parameters.

$$\begin{aligned} lnL(\beta_2, \ldots, \beta_J) &= \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} ln(Pr(y_i = j)) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} ln(\frac{exp(X_i \beta_j)}{\sum_{m=1}^{J} exp(X_i \beta_m)}) \end{aligned}$$

Deriving the MNL in a Random Utility Framework

- Let U_{ij} be the utility of decision-maker i derived when choosing alternative $j \in \{1, \ldots, J\}$ (known to the decision-maker but not to the analyst)
- The decision-maker i chooses alternative with the highest utility, thus alternative j is chosen, if and only if $U_{ij} > U_{ik}$ for all $k \neq j$.
- Given that we cannot observe the decision-makers utility we need to specify a function, such as $V_{ij} = X_i\beta_j$ (the *systematic component* of the model) that relates observed factors X_i, e.g., characteristics of the decision-maker i, to the utility U_{ij}.

Deriving the MNL in a Random Utility Framework

- The *stochastic component* of the model ϵ_{ij} is assumed to be iid extreme value and reflects those parts of the decision-makers utility function that are not covered by V_{ij}. Thus, $U_{ij} = V_{ij} + \epsilon_{ij}$
  - The key assumption here is *independence*. It means that the error for one alternative should provide no information about the error of another alternative.
  - Essentially like assuming that you have a well-specified model.
- If we now consider the difference in utility relative to a common baseline (say, $j = 1$ for identification), for all alternatives in the choice-set, we get:

$$\begin{aligned} \tilde{U}_{ij} &= U_{ij} - U_{i1} \\ &= V_{ij} - V_{i1} + (\epsilon_{ij} - \epsilon_{i1}) \\ &= \tilde{V}_{ij} + \tilde{\epsilon}_{ij} \end{aligned}$$

  Only differences in utility matter!
- Finally, with $\tilde{V}_{ij} = X_i\tilde{\beta}_j$ we get the multinomial logit model

$$Pr(y_i = j|X_i) = \frac{exp(\tilde{V}_{ij})}{\sum_{m=1}^{J} exp(\tilde{V}_{ij})}, \text{ whereby } \tilde{V}_{i1} = X_i\tilde{\beta}_1 = 0$$

MNL - Chooser-specific Data

Suppose we are interested in explaining vote-choice in *Absurdistan* with its stable 3-party-system consisting of $1 =$ the Blues, $2 =$ the Reds and $3 =$ the Greens.

- Suppose the *systematic component* of our model consists of chooser characteristics *Age* and *Education*. Thus we assume to have a well-specified model that estimates effects for those variables that vary across the three parties.
- More clearly, suppose the *systematic component* looks like this:

$$V_{ij} = X_i \beta_j = \beta_{j1} + \beta_{j2} \cdot Age_i + \beta_{j3} \cdot Education_i$$

- Given that the choice-set consists of three parties, the systematic components are:

$$\begin{aligned} V_{i1} &= \beta_{11} + \beta_{12} \cdot Age_i + \beta_{13} \cdot Education_i \\ V_{i2} &= \beta_{21} + \beta_{22} \cdot Age_i + \beta_{23} \cdot Education_i \\ V_{i3} &= \beta_{31} + \beta_{32} \cdot Age_i + \beta_{33} \cdot Education_i \end{aligned}$$

- Thus, $Pr(y_i = j|X_i) = \frac{exp(V_{ij})}{\sum_{j=1}^{3} exp(V_{ij})} = \frac{exp(X_i \beta_j)}{\sum_{j=1}^{3} exp(X_i \beta_j)}$, with $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})' = 0$

MNL needs data in wide-format

RespID	Vote	Age	Education
1	2	18	1
2	1	25	2
3	1	27	2
4	3	43	3

A Table of MNL Estimation Results

TABLE 1 *Disentangling Regime from Anchoring Preferences: A MNL Vote Choice Model of the French Parliamentary Election, 2002*

	Dependent Variable: Vote Choice					
	Others vs. Right			Left vs. Right		
Independent variables	Coef.	Std. err.	p	Coef.	Std. err.	p
Ideology	− 0.580	0.797	0.466	− 2.775	0.823	0.001
Candidate Evaluation	− 2.253	1.033	0.029	− 4.260	1.160	0.000
Right Preference	− 4.463	0.767	0.000	− 3.607	0.963	0.000
Left Preference	− 2.039	0.886	0.021	1.860	0.933	0.046
Unanchored Preference	− 3.174	0.853	0.000	− 1.342	0.931	0.150
Regime-Right	0.318	0.523	0.542	1.962	0.777	0.012
Regime-Left	− 0.812	0.648	0.210	− 1.466	0.591	0.013
Regime-Unanchored	− 0.286	0.622	0.646	− 0.026	0.632	0.968
Economy	0.218	0.359	0.544	− 0.210	0.356	0.554
Security	− 0.255	0.326	0.435	− 0.256	0.415	0.537
Constant	4.571	0.907	0.000	4.678	0.976	0.000

Note: p-values are for two-tailed tests based on robust (White–Huber) standard errors. $N = 670$, and 80 per cent have been correctly classified.

Interpretation of the MNL Estimation Results

Calculate Quantities of Interest

Take a look at the code we used for logit/probit. We just have a different systematic components (including the normalization). Thus, ...

- predicted probabilities.
- first differences.
- ...

You can estimate this model in `R` using `Zelig` or directly using `library(mlogit)`.

Ternary Diagram aka Triplot