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

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova

Week 8 - 6 April 2022

Leftovers from last week

Construction of the Scenario of Interest

There are two conceptual approaches: *Average-Case* vs *Observed Value*

1. So far we have seen examples of the so-called *average-case* approach to define scenario of interest.
    - Set key IV to interesting value, and the other to average (typical) values.
    - Note that those scenarios are hypothetical — they might not actually exist in your data (inference is model dependent!)
    - Nevertheless helpful for theory tests because we often do not hypothesize about an “average case” but about specific conditions under which a relationship holds.
2. The so-called *observed-value* approach is an alternative
    - Provides average marginal effect (AME), i.e. average of the individual marginal effects of the expected change in the probability for a unit change in the key variable of interest
    - Used to calculate average treatment effect (ATE), the average of first-difference when unit gets (hyp.) treatment and when it is not, or more generally the treatment effect of a randomly picked observation.
    - But mean (average) of the marginal effects (AME) $\neq$ marginal effect at the mean (MEMs).

It’s problematic to calculate counterfactuals that are “far away from the data”.

Simulating Parameters - Some Useful Tips

1. Simulate all parameters including ancillary parameters together.
2. Reparameterize to unbounded scale to ...
   - ...make $\tilde{\gamma}$ converge more quickly in n to a multivariate normal. Thus, more reasonable for smaller sample size.
   - ...make maximization algorithm work faster and without explicit constraints.
3. Ideally, all parameters should be unbounded and logically symmetric, e.g.,
   - for non-negative parameters: $\sigma^2 = e^\eta$
   - for $0 - 1$ bounded parameters (e.g., probability): $\pi = \frac{1}{1+e^{-\eta}}$.
   - for $-1 \leq \rho \leq 1$, use $\rho = (e^{2\eta} - 1)/(e^{2\eta} + 1)$ (*Fisher's Z transformation*)

   In all three cases, η is unbounded. Reparameterize back to a scale people care about.

Simulating Parameters - More Useful Tips

- Always compute simulations of Y (i.e. predicted values) and use that as a basis for simulating other quantities (of interest).
- If you are interested in simulating functions of Y, say $ln(Y)$, do the following: Simulate $ln(Y)$ and then apply the inverse function $exp(ln(Y))$ to get Y. Y is probably on a meaningful scale we care about.
- Check *approximation error* of your simulation: Run it twice, check no. of digits of precision that do not change. If *not* enough precision for presenting results in a table, increase M (or m) and try again.
- Analytical calculations (e.g., delta method) and other tricks can speed-up simulation and increase precision.
- `Zelig` does support various models (including Bayesian!) in R.

Introduction

Infrastructure of "Advanced Quantitative Methods" Course

1. Formulate a suitable probability model of the data-generating process including assumptions of how Y is distributed (i.e., stochastic component) and a parameterization of stuff that gets estimated (i.e., systematic component).
2. Write down the (log-)likelihood function based on your parameterization and assumptions.
3. Maximize the log-likelihood, analytically (often hard, even impossible) or numerically (use functions in R).
4. Interpretation of estimation results through simulating quantities of interest.
5. Check whether the assumed model does fit the data.

- You now can apply this infrastructure to any existing model, fine-tune it or come-up with a new one
- Thus, there is not much more to learn for this semester.
- We will learn more models through which you get more familiar with applying this infrastructure.

What should you take home from this class today?

- We expand our toolbox and will learn how we could model ordinal outcomes.
- You get my semi-random thoughts on how to write publishable papers.

Ordered Choice Models

Modeling Ordinal Variables

- Suppose you wanna model a categorical variable.
  Examples: Variables with ordered response categories
    - Likert scales: (1) Almost always, (2) Often, (3) Sometimes, (4) Seldom, (5) Never
    - Agree-Disagree-Items: Strongly Agree, Agree, Disagree, Strongly Disagree
    - Like-Dislike- or Feeling-Thermometer-Scales (e.g., $-5, \ldots, +5$)
    - Classification schemes for events or objects (e.g., No Conflict, Compromise, Conflict)
    - Other ideas from your substantive fields?
- Note that we do not necessarily assume that the differences between the response categories are equal, i.e., (5) - (4) $\neq$ (2) - (1)

Why not using OLS or Logit/Probit instead?

- Dichotomizing variables throws away information (loss of efficiency). Combining adjacent categories does not bias estimates, though. This can be done if ...
    - ...if thresholds cannot be distinguished from one another
    - ...the number of observations in some response categories is small
    - ...or you run into convergence problems during estimation
- If there are many response categories and categories are equidistant OLS is fine; potentially choose a parametrization to avoid out-of-bounds predictions (for instance?). Otherwise you get (as in the case of logit/probit) heteroskedastic errors.

Model Formulation

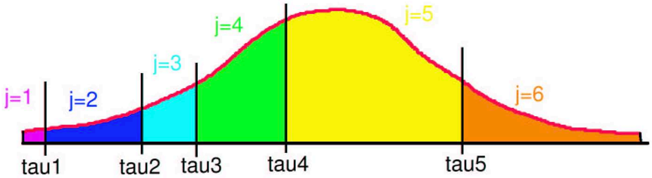
- Let Y^* be a continuous unobserved variable
- Define a model through its stochastic and systematic component

$$
\begin{aligned}
Y_i^* &\sim P(y_i^*|\mu_i) \\
\mu_i &= X_i\beta
\end{aligned}
$$

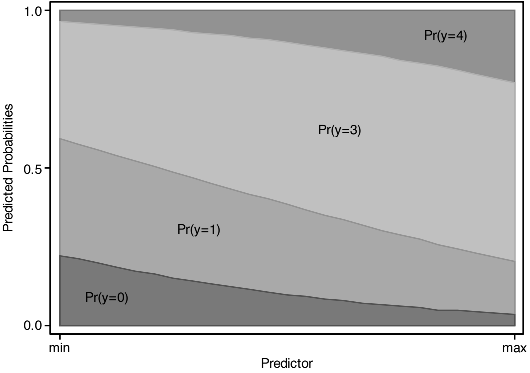
with an *observation mechanism*:

$$
y_{ji} = \begin{cases} 1 & \tau_{j-1} \leq y_i^* < \tau_j \\ 0 & otherwise \end{cases}
$$

- Finally, lets assume independent realizations.



The probabilities across categories sum to 1 for every i



Note, the stochastic component of the model is still Bernoulli:

$$Pr(Y_{ji}|\pi) = \pi_{1i}^{y_{1i}} \cdot \pi_{2i}^{y_{2i}} \cdots \pi_{Ji}^{y_{Ji}}$$

where $\sum_{j=1}^{J} \pi_{ji} = 1$.

Deriving the Likelihood Function

Given the model assumptions, we get the probability of the i^{th} case to be observed in the j^{th} category ($1 \leq j \leq J$) is

$$\begin{aligned}
Pr(Y_i = j) = Pr(Y_{ji} = 1) &= Pr(\tau_{j-1} \leq y_i^* < \tau_j) \\
&= Pr(\tau_{j-1} \leq X_i\beta + \epsilon < \tau_j) \\
&= \int_{\tau_{j-1}}^{\tau_j} P(y_i^*|\mu_i) \\
&= F(\tau_j|\mu_i) - F(\tau_{j-1}|\mu_i) \\
&= F(\tau_j|X_i\beta) - F(\tau_{j-1}|X_i\beta) \\
&= F(\tau_j - X_i\beta) - F(\tau_{j-1} - X_i\beta)
\end{aligned}$$

where F is a cumulative distribution. Consequently, the probability of the i^{th} case to be observed in the ...

...lowest category is $Pr(Y_i = 1) = Pr(Y_{1i} = 1) = F(\tau_1 - X_i\beta)$

...highest category is $Pr(Y_i = J) = Pr(Y_{Ji} = 1) = 1 - F(\tau_{J-1} - X_i\beta)$.

Deriving the Likelihood Function

Given that each observation falls within one and only one category 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 $L_i(\beta, \tau)$ of observation i is

$$\begin{aligned} lnL_i(\beta,\tau) &= ln(\prod_{j=1}^{J} Pr(Y_{ji}=1)^{y_{ji}}) \\ &= \sum_{j=1}^{J} y_{ji} ln(Pr(Y_{ji}=1)) \end{aligned}$$

Deriving the Likelihood Function

Then summing-up all individual contributions assuming independent realizations gives us the log-likelihood of this ordered choice model.

$$
\begin{aligned}
lnL(\beta,\tau) &= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ji} ln(Pr(Y_{ji}=1)) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ji} ln(F(\tau_j - X_i\beta) - F(\tau_{j-1} - X_i\beta)) \\
&= \sum_{y_i=1} ln(F(\tau_1 - X_i\beta)) \\
&\vdots \\
&+ \sum_{y_i=j} ln(F(\tau_j - X_i\beta) - F(\tau_{j-1} - X_i\beta)) \\
&\vdots \\
&+ \sum_{y_i=J} ln(1 - F(\tau_{J-1} - X_i\beta))
\end{aligned}
$$

Model Formulation

- Choose $\Phi(\cdot)$ for $F(\cdot)$, i.e. assuming standard normal distributed errors of the latent regression on y^*, we get the *ordered probit model*.
- Choose $\Lambda(\cdot)$ for $F(\cdot)$, i.e. assuming a standard logistic errors of the latent regression on y^*, we get the *ordered logit model*.

Model Identification

- We have k parameters in β and $J-1$ thresholds in τ.
- We cannot estimate all parameters, because

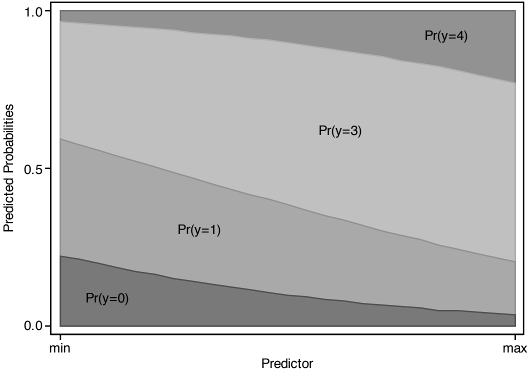
$$Pr(y_i = 1) = F(\tau_1 - X_i\beta) = F(\tau_1 - \beta_0 - X_{i*}\beta_*)$$

  and if we add a constant c to both, threshold and intercept we get exactly the same terms.

$$\begin{aligned} F((\tau_1 + c) - (\beta_0 + c) - X_{i*}\beta_*) &= F(\tau_1 + c - \beta_0 - c - X_{i*}\beta_*) \\ &= F(\tau_1 - \beta_0 - X_{i*}\beta_*) \\ &= Pr(y_i = 1) \end{aligned}$$

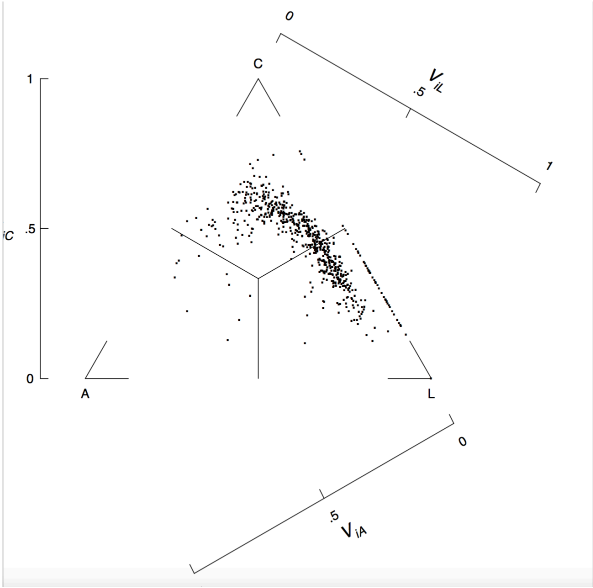
  Thus, the model with separate parameters for $J-1$ thresholds and intercept β_0 is not identified.
- However, the model can be identified by fixing the variance and ...
  - ...removing the constant and estimating all τ_j (`Zelig` does this!)
  - ...setting $\tau_0 = 0$ and estimating the constant
- Take a look at the Jackman piece for further instructive ideas (e.g. fix two thresholds)

Interpretation



- Model predictions are J probabilities that sum to 1.
- One first difference (i.e., change of scenario in one independent variable) has an effect on all J probabilities.
- When one probability goes up, at least one of the others must go down.

Use tenary diagram aka triplot if $J = 3$



Ordered Probit implementation in R

```
## Make a matrix nrows = no. observations, ncols = no. categories
## with 1's or 0's if that observation is in that category (y_ji)
y0 <- sort(unique(y))    # get categories of the DV
m = length(unique(y))    # no of categories of the DV
Y_ji <- matrix(NA, nrow=length(y), m)
for (j in 1:m){
    Y_ji[,j] <- y==y0[j]
}

##Code up the loglikelihood function
ll.oprobit <- function(theta, Y_ji, X){
    m<-dim(Y_ji)[[2]]                          # number of categories
    beta <- theta[1:ncol(X)]
    tau<-theta[(ncol(X)+1):(ncol(X)+(m-1))]     # tau is of length m-1
    ystarmu <- X%*%beta
    probs <- matrix(nrow=nrow(ystarmu), ncol=m)

    probs[,1] <- pnorm(tau[1]-ystarmu)          # 1st category

    for (j in 2:(m-1))                          # categories inbetween
    probs[,j] <- pnorm(tau[j]-ystarmu) - pnorm(tau[j-1]-ystarmu)

    probs[,m] <- 1-pnorm(tau[m-1]-ystarmu)      # last category

    sum(log(probs[Y_ji]))
    }
```

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 think 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.