UNIVERSITY OF MANNHEIM
School of Social Sciences

Advanced Quantitative Methods in Political Science: Multi-Level Models

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova

Week 12 - 18 May 2022

What should you take home from this class today?

- We will finish-up some left-overs from last week regarding Heckman models.
- We will meet *multi-level models* as a way to deal with multi-level data structures.

The Heckman Model & Selection Bias – some leftovers

- Let Y_i vector for observation i $(= 1, \dots, n)$
- $Y_i = (Y_{1i}^*, Y_{2i})'$ is bi-variate normal distributed with an
  1. *Selection equation*:
  $$y_{1i}^* = \mu_{1i} + u_i = X_{1i}\beta_1 + u_i, \;\; u_i \sim N(0, 1)$$
  with an *stochastic censoring mechanism* (a.k.a *sample selection rule*)
  $$y_{1i} = \begin{cases} 1 & y_{1i}^* > 0 \\ 0 & y_{1i}^* \leq 0 \end{cases}$$
  2. *Outcome equation*: For all selected observations i, i.e., if $y_{1i}^* > 0$ one has
  $$y_{2i} = \mu_{2i} + \epsilon_i = X_{2i}\beta_2 + \epsilon_i, \;\; \epsilon_i \sim N(0, \sigma_2^2)$$
  whereby the error terms of both equations are correlated, i.e. $0 \neq \rho = corr(u_i, \epsilon_i)$. (Note that we get a Tobit as a special case if $y_{1i}^* = y_{2i}$)
- Thus, (check the dimensionality!)
$$Y_i = \begin{pmatrix} Y_{1i}^* \\ Y_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right]$$

.... and the Likelihood

The likelihood function is a combination of the likelihood for censored and uncensored observations

- For $y_{1i}^* \leq 0$ all that is observed is that this event occurred. Thus, the density is the probability $Pr(y_{1i}^* \leq 0)$ that it occurred
- For $y_{1i}^* > 0$ we observe y_{2i} with a certain (conditional) probability. It is the probability of being selected, $Pr(y_{1i}^* > 0)$, multiplied by the bivariate density $f(y_{2i}|y_{1i}^* > 0)$.
- The likelihood function of a bivariate sample selection model is as follows:

$$L(\beta_1, \beta_2, \rho, \sigma_2^2) = \prod_{i=1}^{n} Pr(y_{1i}^* \leq 0)^{1-y_{1i}} \cdot \{f(y_{2i}|y_{1i}^* > 0) \cdot Pr(y_{1i}^* > 0)\}^{y_{1i}}$$

- One can show that the second term simplifies to a univariate normal distribution that can be easily handled computationally. Details can be found, for instance, in Amemiya's *Advanced Econometrics* (1985: 385-7) textbook.

Application: A Model to predict success in the Graduate Program

- Suppose we like to test whether GRE-scores predict success in terms of grades in our PhD-program.
- Obviously, in all the application files we have GRE-scores. Grades, however, are only available for the ones who join our program.
- Thus, the bottom of the distribution of the unobserved variable (Y^*_{1i}), *Admission Rating* of our PhD-program, is censored. The PhD selection committee only admits those candidates and monitor their performance who rate high on the latent *Admission Rating* variable.

Application: A Model to predict success in the Graduate Program

- Suppose we have the following *selection* and *outcome equation*:
  1. *Selection equation*:

$$AdmissionRating = \beta_{10} + \beta_{11}GRE + \beta_{12}TOEFL + u_i, \ u_i \sim N(0,1)$$

     with a *censoring mechanism* (a.k.a *sample selection rule*)

$$Admission = \begin{cases} 1 & AdmissionRating > 0 \\ 0 & AdmissionRating \leq 0 \end{cases}$$

  2. *Outcome equation*: For all enrolled (and former) students i, i.e., if $AdmissionRating > 0$ one has

$$Success = \beta_{20} + \beta_{21}GRE + \beta_{22}Math + \epsilon_i, \ \epsilon_i \sim N(0, \sigma_2^2)$$

- Admitted graduate students are not representative of applicants generally. Despite low GRE-scores applicants get admitted if they have high TOEFL-scores or because they have large error term – i.e., their applications have qualities that are uncorrelated with GRE or TOEFL scores (e.g., strong letter, University's reputation).

What if we just run an OLS for all observations we have data for?

- Group of students that were admitted because of high GRE scores are representative of the group of applicants with high GRE scores.
- However, the group of admitted students with low GRE scores are *not* representative of the group of all applicants with this score. Assuming that the selection committee has done a good job, those admitted low-score students perform better than the non-admitted ones.
- Thus, running regression on the selected sample might wrongly show that GRE does not systematically predict success in graduate school.

Identification, Interpretation and Estimation

- Selection bias models, as the Timpone-Example shows, are not bounded to have a normally distributed stochastic component but can be in principle fit to any theory about the selection and outcome processes.
- *Identification* of those processes is an issue with selection bias models, though. You need at least one variable (and the more the better!) that only predicts *selection* but not the *outcome* (otherwise identification hinges solely on non-linearity of the selection equation, hence on distributional assumptions that cannot be checked rigorously).
- *Interpretation.* As usual, calculate expected values, predicted probabilities and first-differences using statistical simulations.
- You can estimate these models in `R` for instance using the `library(sampleSelection)`.

Case Selection and Selection Bias

Selection on the Dependent Variable

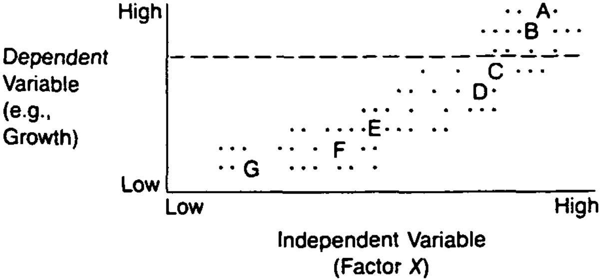


Fig. 1. Assumed relationship between factor *X* and the dependent variable

Taken from: Geddes, Barbara. 1997. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics" *Political Analysis* 2(1): 131-50; Figure 1.

Selection on the Dependent Variable

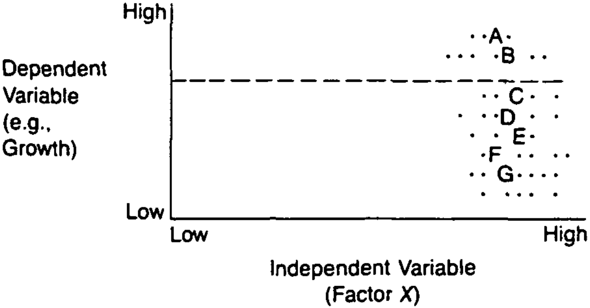


Fig. 2. An alternative possibility for the relationship between factor *X* and the dependent variable

Taken from: Geddes, Barbara. 1997. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics" *Political Analysis* 2(1): 131-50; Figure 2.

Selection on the DV - Endpoints in a Time Series

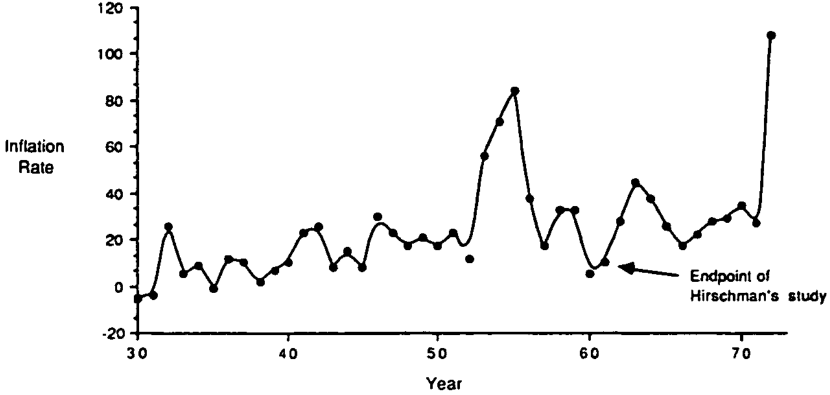


Fig. 11. Inflation in Chile, 1930–72. (Data for 1930–61 from Hirschman

Taken from: Geddes, Barbara. 1997."How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics" *Political Analysis* 2(1): 131-50; Figure 11.

Quiz

Quiz

You are a reviewer of a manuscript in which the author analyzes the US president's use of force in the conduct of American foreign policy depending on the president's popularity in the polls. The author coded each presidential response on a scale ranged from 0 (compliance with an opponent's demands) to 10 (violent military action) with nine intermediate responses unequally spaced between 0 and 10. "Because the dependent variable is bounded below by 0," the author analyzes these data with a Tobit model.

What would you write in your report to the journal editor?

1. I'd reject the paper because no censoring was involved. 0 is not a value at which certain naturally occurring values had been censored.
2. I would suggest a more appropriate model that restricted the range of y without assuming censoring.
3. I would suggest to recode the dependent variables to be equally spaced in order to avoid biased estimates.
4. I accept the paper because the author rightly chooses a Tobit model.

Multi-Level Models

Multilevel Data Structures

- Data is structured hierarchically
- Units of analysis are a subset of other units for which data is also available
- Familiar substantive hierarchies are ...
    - ...voters nested in districts (nested in regions)
    - ...individual nested in associations (nested in countries)
    - ...countries nested in organizations
    - ...parties nested in governments
    - ...others?
- Conceptual hierarchies are ...
    - ...time nested in states (e.g., years in countries)
    - ...longitudinal data: waves nested within individuals (panel)
    - ...measurement models: measurements nested within individuals (e.g., indicators of a latent factor, Item – Response – Theory (IRT) models)
- "Once you know that hierarchies exist, you see them everywhere" (Kreft and de Leeuw. 1998. *Introducing Multilevel Modeling*, p. 1)

Basic Idea

- Explicit model for this data structure (instead of statistical fix)
- Model consists of modeled and unmodeled coefficients.
- Each level of analysis has its own regression model, with different assumptions about error distribution, functional form ect.
- Ideal data situation to apply multilevel models:
  data on many group-level (level-2) observations (Stegmueller, 2014 suggests $J > 20$),
  no. of unit-level (level-1) observations can be sparse, though
- We distinguish three different scenarios:
  (1) Varying Intercept Model
  (2) Varying Slope Model
  (3) Varying Intercept - Varying Slope Model

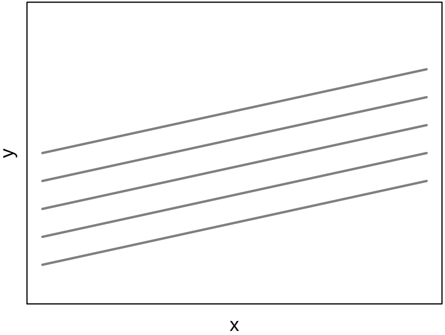
(1) Varying Intercept Model

- Basic linear model with common slope for all j groups but different (aka varying) intercepts

$$y_{ij} = \beta_{0j[i]} + \beta_1 x_i + \epsilon_{ij}$$

or equivalently as $y_i \sim N(\beta_{0j[i]} + \beta_1 x_i,\ \sigma_y^2)$

- $j[i]$ indicates that case i gets a group-specific intercept j

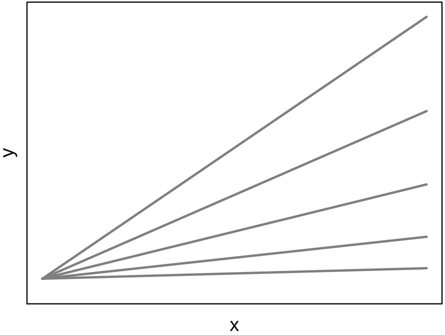


(2) Varying Slope Model

- Common intercept for all j groups but different (aka varying) slopes

$$y_{ij} = \beta_0 + \beta_{1j[i]}x_i + \epsilon_{ij}$$

  or equivalently as $y_i \sim N(\beta_0 + \beta_{1j[i]}x_i,\ \sigma_y^2)$
- $j[i]$ indicates that case i gets a group-specific slope j

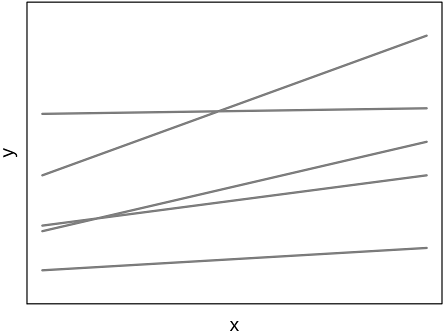


(3) Varying Intercept - Varying Slope Model

- Group-specific intercepts as well as group-specific slopes

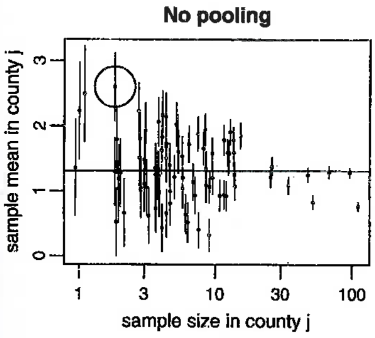
$$y_{ij} = \beta_{0j[i]} + \beta_{1j[i]}x_i + \epsilon_{ij}$$

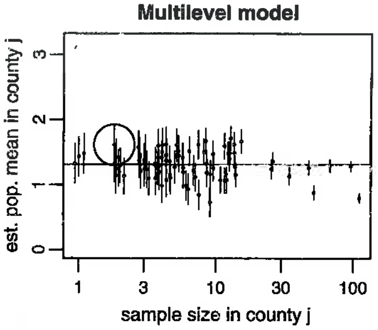
  or equivalently as $y_i \sim N(\beta_{0j[i]} + \beta_{1j[i]}x_i,\ \sigma_y^2)$
- $j[i]$ indicates that case i gets a group-specific intercept *and* slope j



Multilevel Models as Partial Pooling Models

- *Complete pooling:*
  - *One model fits all, irrespective of group membership*
  - *Variation between groups is ignored (although potentially relevant)*
- *No pooling:*
  - *Separate models for each group*
  - *Potentially to much emphasis on group differences; they look more different than they are*
- *Partial pooling:*
  - *Captures similarity between groups and uniqueness within groups*
  - *Multilevel models are a compromise between no pooling and complete pooling*


No pooling
sample mean in county j
0
1
2
3
1
3
10
30
100
sample size in county j


Multilevel model
est. pop. mean in county j
0
1
2
3
1
3
10
30
100
sample size in county j

A Simple Multilevel Model

- Lets start by assuming the following *level-1 (unit-level) model*

$$y_{ij} = \beta_{0j[i]} + \beta_{1j[i]} x_i + \epsilon_{ij}$$

- Now lets model the variation of level-1 regression parameters as a function of level-2 variable z, the so-called *level-2 (group-level) model*

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + \delta_{0j} \text{ and } \beta_{1j} = \gamma_{10} + \gamma_{11} z_j + \delta_{1j}$$

- Although the model is fully characterized (given the error-term assumptions) we can substitute in the level-2 model into the level-1 model to derive a single-equation expression

$$\begin{aligned} y_{ij} &= (\gamma_{00} + \gamma_{01} z_j + \delta_{0j}) + (\gamma_{10} + \gamma_{11} z_j + \delta_{1j}) x_i + \epsilon_{ij} \\ &= \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_i + \gamma_{11} z_j x_i + \delta_{0j} + \delta_{1j} x_i + \epsilon_{ij} \end{aligned}$$

- Thus, this is a regression model with a 4 parameters (constant, level-1 and level-2 effect, cross-level interaction effect) as well as 3 different error-terms (that we try to identify separately)

A Simple Multilevel Model

The model so far is incomplete without specifying the assumptions about the error terms

- Level-1 error term: $\epsilon_{ij} \sim N(0, \sigma_y^2)$.
  Of course, if the dependent variable stems from a count, duration or binomial process, other assumptions have to be evoked.
- Level-2 error terms are assumed to follow the following bivariate normal distribution:

$$\begin{pmatrix} \delta_{0j} \\ \delta_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\beta_0} & \rho\, \sigma_{\beta_0}\sigma_{\beta_1} \\ \rho\, \sigma_{\beta_0}\sigma_{\beta_1} & \sigma^2_{\beta_1} \end{pmatrix} \right] ;\ j = 1, \ldots, J$$

- Finally, we need to assume that errors at both levels are uncorrelated, i.e. $Cov[\delta_{0j}, \epsilon_{ij}] = Cov[\delta_{1j}, \epsilon_{ij}] = 0$.
- Moreover, one can show that the total error variance is $Var(\delta_{0j} + \delta_{1j}x_i + \epsilon_{ij}) = \sigma^2_{\beta_0} + 2x_i\rho\, \sigma_{\beta_0}\sigma_{\beta_1} + x_i^2\sigma^2_{\beta_1} + \sigma^2_y$. Thus, we get a non-constant variance (hence, heterogeneity), although ϵ_{ij}, δ_{0j} and δ_{1j} have constant variance.

Describing Variance Partitioning

- A useful summary statistic is the variance partitioning coefficient (aka *intra-class correlation*), given that $\delta_{1j} = 0$ (i.e., no varying slope)

$$ICC = \frac{\sigma^2_{\beta_0}}{\sigma^2_{\beta_0} + \sigma^2_y} = \frac{group-level\ variation}{total\ variation}$$

- Interpretation:
  - Correlation of two randomly drawn units of a group.
  - Proportion of the total variance contributed by the grouping variance component.
  - ICC varies between 0 and 1. When $ICC = 0$, the grouping variance is unimportant and the multi-level estimator is no different from the pooled estimator.
  - Use LRT to test this (because the pooled model is nested within the multi-level model)

Application - EU support across Countries (Steenbergen & Jones, 2002)

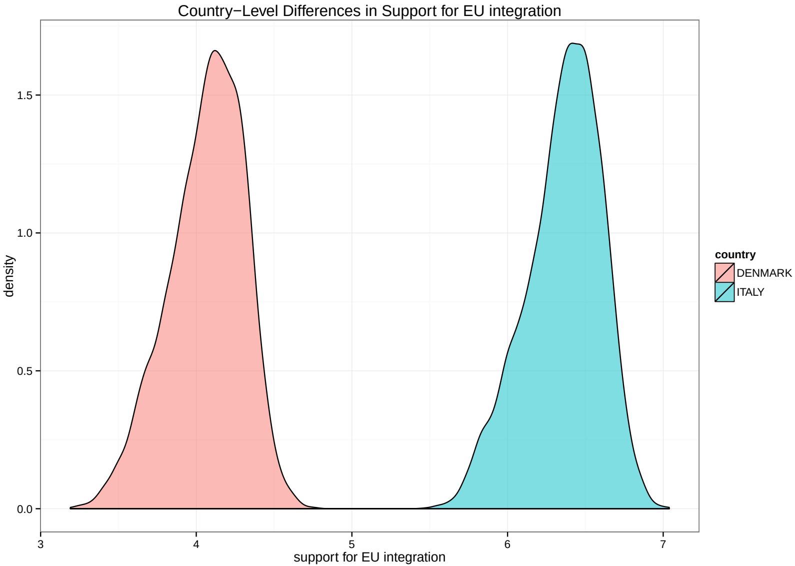
- Individual-level factors (e.g, Ideology, Age, Income, Gender, Opinion Leadership scale): $support_{ijk} = \alpha_{0jk} + X_{ijk}\alpha_{jk} + \epsilon_{ijk}$
- Party-level factors (e.g., Elite perception of party position on EU Integration): $\alpha_{0jk} = \beta_{00k} + Z_{jk}\beta + \delta_{0jk}$
- Country-level factors (e.g., length of EU membership, trade): $\beta_{00k} = \gamma_{000} + T_k\gamma + \nu_{00k}$
- Include cross-level interaction: Effect of *Opinion Leadership* (α_{*jk}) should vary as a function of *party cue* (i.e., we account for *causal heterogeneity*). Thus, $\alpha_{*jk} = \beta_{*0k} + \beta_{*1k}Cue_{jk} + \delta_{*jk}$

The Ugly Table

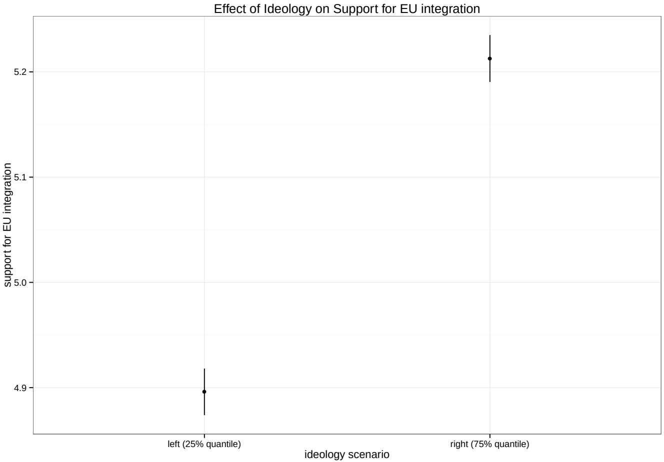
Parameter	Estimate
Fixed Effects	
Constant	5.507**
	(.220)
Tenure	0.015
	(.014)
Trade	0.031
	(.025)
Partisan Cue	0.225**
	(.028)
Lowest Income Quartile	–.104+
	(.064)
Highest Income Quartile	0.046
	(.058)
Ideology	0.019
	(.015)
Opinion Leadership	0.148**
	(.034)
Male	0.088+
	(.050)
Age	–.013**
	(.002)
Opinion Leadership × Partisan Cue	0.044*
	(.021)
Variance Components	
Country-Level (ω_{00})	0.548**
	(.212)
Party-Level	
Constant (τ_{00})	0.107**
	(.031)
Opinion Leadership (τ_{44})	0.019
	(.012)
Constant, Opinion Leadership (τ_{04})	–.011
	(.015)
Individual-Level (σ^2)	3.751**
	(.067)
–2 × Log Likelihood	26578.970

Taken from: Steenbergen and Jones. 2002 - “Modeling Multilevel Data Structures”, *American Journal of Political Science 46(1): 218–237; Table 5.*

Quantities of Interest



Quantities of Interest



Support for EU integration is measured on 0 – 8 scale.