rcrologit

The package provides estimation and inferential procedures for rank-ordered logit model with agents with heterogeneous taste preferences.

Setup

We have n i.i.d. random draws

$$\mathcal{D} := (Y_i, X_i)_{i=1}^n$$

where $Y_i := (Y_{i0}, Y_{i1}, \dots, Y_{iJ})^\top, Y_{i\ell} = 0, 1, \dots, J$, is a vector of ranks and $C_i := (X_{i0}, C_{i1}, \dots, C_{iJ})^\top \in \mathbb{R}^{(J+1)\cdot K}$, $C_{i\ell} \in \mathbb{R}^K$. Let the latent utility model (McFadden, 1974) be

$$U_{ij}^\star = u_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{\texttt{iid}}{\sim} \mathsf{Gu}(0,1).$$

For notational convenience, define the functions $r_i : \{0, 1, \dots, J\} \to \{0, 1, \dots, J\}, i = 1, \dots, n$. Such functions map the rank $j \in \{0, 1, \dots, J\}$ into the corresponding item $r(j) \in \{0, 1, \dots, J\}$ according to individual i's preferences. To clarify

$$Y_{ij} = k \quad \Longleftrightarrow \quad r_i(k) = j$$

An observed ranking for a respondent implies a complete ordering of the underlying utilities. An individual will prefer an item with a higher utility over an item with a lower utility. If we observe a full ranking $r_i := (r_i(0), r_i(1), \dots, r_i(J))^\top$, we know that

$$U_{ir_i(0)}^\star > U_{ir_i(1)}^\star > \cdots > U_{ir_i(J)}^\star.$$

Therefore, the probability of observing a particular ranking r_i is given by

$$\mathbb{P}\left[r_i \mid \mathcal{D}\right] = \mathbb{P}\left[U_{ir_i(0)}^\star > U_{ir_i(1)}^\star > \cdots > U_{ir_i(J)}^\star \mid \mathcal{D}\right] = \prod_{j=0}^{J-1} \frac{\exp\left(u_{ir_i(j)}\right)}{\sum\limits_{j \le \ell \le J} \exp\left(u_{ir_i(\ell)}\right)}.$$

In light of this, we can see that the rank-ordered logit is nothing else than a series of multinomial logit (MNL) models: when $j = 0$ we considered a MNL the most preferred item; another MNL for the second-ranked item to be preferred over all items except the one with rank 1, and so on. Finally, the probability of a complete ranking is made up of the product of these separate MNL probabilities. The product contains only J probabilities, because ranking the least preferred item is done with probability 1.

Modelling

In its most general form, we allow the user to model $u_{i\ell}$ in the latent utility model as

$$u_{i\ell} = X_{i\ell}^\top \boldsymbol{\beta}_{\texttt{F}} + Z_i^\top \boldsymbol{\alpha}_{\ell,\texttt{F}} + W_{i\ell}^\top \boldsymbol{\beta}_i + V_i^\top \boldsymbol{\alpha}_{i\ell} + \delta_\ell$$

An alternative, handier way to rewrite the model above is to define $Z_{i\ell} := \sum\limits_{j=1}^J Z_i \times \mathbf{1}(j = \ell)$ and $V_{i\ell} := \sum\limits_{j=1}^J V_i \times \mathbf{1}(j = \ell), \ell = 1, 2, \dots, J$, and consider the equivalent model

$$u_{i\ell} = X_{i\ell}^\top \boldsymbol{\beta}_{\texttt{F}} + Z_{i\ell}^\top \boldsymbol{\alpha}_{\texttt{F}} + W_{i\ell}^\top \boldsymbol{\beta}_i + V_{i\ell}^\top \boldsymbol{\alpha}_i + \delta_\ell,$$

where:

- $X_{i\ell}$ are covariates varying at the unit-alternative level whose coefficients are modelled as fixed
- Z_i are covariates varying at the unit level whose coefficients are modelled as fixed
- $W_{i\ell}$ are covariates varying at the unit-alternative level whose coefficients are modelled as random
- V_i are covariates varying at the unit level whose coefficients are modelled as random the random coefficients
- The heterogeneous taste coefficients are modeled as a joint multivariate normal and are i.i.d. across units with mean $\left[\boldsymbol{\alpha}_{\mathtt{R}}{}^\top, \boldsymbol{\beta}_{\mathtt{R}}{}^\top\right]^\top$ and variance $\boldsymbol{\Sigma}$.
- δ_ℓ are alternative-specific fixed effects
- $\epsilon_{i\ell} \sim \mathsf{Gu}(0,1)$ are idiosyncratic i.i.d. shocks

Note that whenever $W_{i\ell}$ and V_i are not specified estimates a standard rank-ordered logit with no heterogeneous preferences and the conditional choice probabilities are given by

$$\mathbb{P}\left[r_i \mid \mathcal{D}\right] = \prod_{j=0}^{J-1} \frac{\exp\left(u_{ir_i(j)}\right)}{\sum\limits_{j\leq\ell\leq J} \exp\left(u_{ir_i(\ell)}\right)}.$$

If instead agents are allowed to have heterogeneous taste, then

$$\mathbb{P}[r_i \mid \mathcal{D}] = \int \prod_{j=0}^{J-1} \frac{\exp\left(u_{ir_i(j)}^\top(\beta_i)\right)}{\sum\limits_{j\leq\ell\leq J} \exp\left(u_{ir_i(\ell)}^\top(\beta_i)\right)} \phi(\beta_i;\beta,\Sigma)\mathrm{d}\beta_i.$$

The parameter vector to be estimated is thus

$$\theta = \left(\boldsymbol{\beta}_{\mathtt{F}}{}^\top, \boldsymbol{\beta}_{\mathtt{R}}{}^\top, \boldsymbol{\alpha}_{\mathtt{F}}{}^\top, \boldsymbol{\alpha}_{\mathtt{R}}{}^\top, \mathrm{vech}(\boldsymbol{\Sigma})^\top, \{\delta\}_{j=1}^J\right)^\top.$$

Estimation

The ideal maximum likelihood estimator is defined as

$$\widehat{\theta}_{\mathtt{ML}} := \arg\max_\theta \sum_{i=1}^n \log \int \prod_{j=0}^{J-1} \frac{\exp\left(u_{ir_i(j)}(\theta)\right)}{\sum\limits_{j\leq\ell\leq J} \exp\left(u_{ir_i(\ell)}(\theta)\right)} \phi(\beta_i;\beta_{\mathtt{R}},\Sigma)\mathrm{d}\beta_i.$$

We approximate the integral via montecarlo as

$$\widehat{\mathbb{P}}_{(\widehat{\beta},\widehat{\Sigma})}[r_i \mid \mathcal{D}] = \frac{1}{S}\sum_{i=1}^S \prod_{j=0}^{J-1} \frac{\exp\left(u_{ir_i(j)}(\theta, \beta_i^{(s)})\right)}{\sum\limits_{j\leq\ell\leq J} \exp\left(u_{ir_i(\ell)}(\theta, \beta_i^{(s)})\right)},$$

where $\boldsymbol{\beta}_i \overset{\mathtt{iid}}{\sim} \mathsf{N}(\widehat{\boldsymbol{\beta}}_{\mathtt{R}}, \widehat{\boldsymbol{\Sigma}})$.

Installation

You can install the development version of rcrologit from GitHub with:

```
# install.packages("devtools")
devtools::install_github("filippopalomba/rcrologit")
```

Basic Usage

```
library(rcrologit)

data <- rcrologit_data

# Rank-ordered logit
dataprep <- dataPrep(data, idVar = "Worker_ID", rankVar = "rank",
                     altVar = "alternative",
                     covsInt.fix = list("Gender"),
                     covs.fix = list("log_Wage"), FE = c("Firm_ID"))

rologitEst <- rcrologit(dataprep)

# Rank-ordered logit
dataprep <- dataPrep(data, idVar = "Worker_ID", rankVar = "rank",
                     altVar = "alternative",
                     covsInt.het = list("Gender"),
                     covs.fix = list("log_Wage"), FE = c("Firm_ID"))

rologitEst <- rcrologit(dataprep, stdErr="skip")
```