# Package ‘rcrologit’

February 15, 2024

**Title** Random Coefficients Rank-Ordered Logit

**Version** 0.0.4

**Description** Estimation and inference procedures for random coefficients rank-ordered logit.

**Depends** R (>= 4.1.0)

**Imports** fastDummies (>= 1.6.3), MASS (>= 7.3), methods (>= 4.2.1), numDeriv (>= 2016.8), parallel (>= 4.2.1), purrr (>= 1.0.1), Rdpack (>= 2.4), rlang (>= 1.1.0), tibble (>= 3.2.0)

**RdMacros** Rdpack

**License** MIT + file LICENSE

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.2.3

**URL** https://github.com/filippopalomba/rcrologit

**BugReports** https://github.com/filippopalomba/rcrologit/issues

**NeedsCompilation** no

**LazyData** true

**Author** Filippo Palomba [aut, cre] (<https://orcid.org/0000-0002-6400-3142>), Chiara Motta [aut]

**Maintainer** Filippo Palomba <fpalomba@princeton.edu>

## R topics documented:

---

`rcrologit-package` *rcrologit: A Package to Estimate Random Coefficients Rank-Ordered Logit Models.*

---

### Description

The package implements estimation and inference procedures for random coefficient rank-ordered logit models.

### Author(s)

Chiara Motta, University of California Berkeley. `<cmotta@berkeley.edu>`

Filippo Palomba, Princeton University. `<fpalomba@princeton.edu>`

### See Also

Useful links:

- `https://github.com/filippopalomba/rcrologit`
- Report bugs at `https://github.com/filippopalomba/rcrologit/issues`

---

`dataPrep` *Data Preparation for Estimation of Random Coefficients Rank-Ordered Logit*

---

### Description

This function prepares the data for estimation of a random coefficient rank-ordered logit. The rank-ordered logit - sometimes termed *exploded logit model* - was originally proposed in Beggs et al. (1981) and it is an extension of the Luce (1959)-McFadden (1974) model.

These models automatically implies independence of irrelevant alternatives (Debreu 1960). However, McFadden and Train (2000) show that if agents are allowed to have heterogeneous tastes (i.e., random coefficients), then the conditional probability choices induced by the latent utility model can match those implied by virtually any discrete choice probability model.

The package `rcrologit`, depending on the type of covariates specified in `dataPrep`, allows the user to estimate:

- the standard rank-ordered logit model if either `covs.fix` or `covsInt.fix` are specified.
- the random coefficients rank-ordered logit model if either `covs.het` or `covsInt.het` are specified.
- the mixed random coefficients rank-ordered logit model if at least one of `covs.fix` or `covsInt.fix` and at least one of `covs.het` or `covsInt.het` are specified.

For more information on the underlying specification see the official repository.

### Usage

```
dataPrep(
  dataraw,
  idVar,
  rankVar,
  altVar,
  FE = NULL,
  covs.fix = NULL,
  covs.het = NULL,
  covsInt.fix = NULL,
  covsInt.het = NULL
)
```

### Arguments

| | |
|---|---|
| `dataraw` | a `data.frame` object containing the data to be prepared for estimation |
| `idVar` | a string indicating the name of the unit identifier variable |
| `rankVar` | a string indicating the name of the variable indicating the rank of each alternative. It should be a variable containing integer values ranging from 1 to $J$, where $J$ is the number of alternatives. Each unit's best alternative must get value 1, the second best value 2, and so on in a decreasing fashion. |
| `altVar` | a string indicating the name of the alternative identifier variable |
| `FE` | a string containing the name of the fixed effect variable (if required by the user) |
| `covs.fix` | covariates varying at the alternative-unit level with fixed taste parameter |
| `covs.het` | covariates varying at the alternative-unit level with random taste parameter |
| `covsInt.fix` | covariates varying at the unit level with fixed taste parameter |
| `covsInt.het` | covariates varying at the unit level with random taste parameter |

### Details

The function prepares the data to estimate a random coefficient rank-ordered logit model induced by the latent utility model McFadden (1974)

$$U_{i\ell} = u_{i\ell} + \epsilon_{i\ell}, \quad i = 1, 2, \ldots, n, \quad j = 0, 1, \ldots, J.$$

In its most general form, we model $u_{i\ell}$ as

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

where

- $X_{i\ell}$ are covariates varying at the unit-alternative level whose coefficients are modelled as fixed. The user can specify these covariates via the option `covs.fix`.
- $Z_i$ are covariates varying at the unit level whose coefficients are modelled as fixed. The user can specify these covariates via the option `covsInt.fix`. These coefficients are interacted with a dummy for the choice and treated as alternative varying covariates, i.e., $Z_{i\ell} = \sum_{j=1}^J Z_i \times \mathbf{1}(j = \ell)$, where $J = 0$ is the reference group.
- $W_{i\ell}$ are covariates varying at the unit-alternative level whose coefficients are modelled as random. The user can specify these covariates via the option `covs.het`.

- $V_i$ are covariates varying at the unit level whose coefficients are modelled as random. The user can specify these covariates via the option `covsInt.het`. These coefficients are interacted with a dummy for the choice and treated as alternative varying covariates, i.e., $V_{i\ell} = \sum_{j=1}^{J} V_i \times \mathbf{1}(j = \ell)$, where $J = 0$ is the reference group.
- the random coefficients are modeled as a joint multivariate normal and are i.i.d. across units,

$$\begin{bmatrix} \boldsymbol{\alpha}_i \\ \boldsymbol{\beta}_i \end{bmatrix} \sim \mathsf{N}\left(\begin{bmatrix} \boldsymbol{\alpha}_{\mathtt{R}} \\ \boldsymbol{\beta}_{\mathtt{R}} \end{bmatrix}, \boldsymbol{\Sigma}\right)$$

- $\delta_\ell$ are alternative-specific fixed effects that can be specified via the option `FE`.
- $\epsilon_{i\ell} \sim \mathsf{Gu}(0, 1)$are idiosyncratic i.i.d. shocks.

The parameter vector to be estimated is thus

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

where the first alternative-fixed effect has been normalized to 0.

For more information on the underlying specification see the official repository.

## Value

The function returns a list containing the following objects:

| | |
|---|---|
| `id` | vector containing the id of each observation |
| `rank` | vector containing the rank of each observation |
| `X.fix` | a matrix containing covariates with fixed taste parameter |
| `X.het` | a matrix containing covariates with random taste parameter |
| `param.spec` | a list containing some parameters describing the specification chosen by the user |

## Author(s)

Chiara Motta, University of California Berkeley. <cmotta@berkeley.edu>

Filippo Palomba, Princeton University. <fpalomba@princeton.edu>

## References

Beggs S, Cardell S, Hausman J (1981). "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics*, **17**(1), 1–19. ISSN 0304-4076, doi:10.1016/03044076(81)900567, `2023-03-09`.

Debreu G (1960). "Review of RD Luce, Individual Choice Behavior: A Theoretical Analysis." *American Economic Review*, **50**(1), 186–188.

Luce RD (1959). *Individual Choice Behavior: A Theoretical Analysis*. Courier Corporation. ISBN 978-0-486-15339-1.

McFadden D (1974). "Frontiers in Econometrics." In *Conditional Logit Analysis of Qualitative Choice Behavior*, 105–142. Academic Press. ISBN 0-12-776150-0.

McFadden D, Train K (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, **15**(5), 447–470. ISSN 1099-1255, doi:10.1002/10991255(200009/10)15:5<447::AID-JAE570>3.0.CO;21, `2023-05-04`.

## See Also

rcrologit

## Examples

```
data <- rcrologit_data
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)
```

---

`print.rcrologit` *Print Method for Random Coefficients Rank-Ordered Logit*

---

## Description

The print method for for random coefficients rank-ordered logit fitted objects.

## Usage

```
## S3 method for class 'rcrologit'
print(x, printFE = FALSE, ...)
```

## Arguments

| | |
|---|---|
| `x` | Class "rcrologit" object, obtained by calling `rcrologit`. |
| `printFE` | Whether fixed effects coefficients and standard errors should be displayed. |
| `...` | Other arguments. |

## Value

No return value, called to print `rcrologit` results.

## Author(s)

Chiara Motta, University of California Berkeley. <cmotta@berkeley.edu>

Filippo Palomba, Princeton University. <fpalomba@princeton.edu>

## See Also

`rcrologit` for estimation and inference for random coefficients rank-ordered logit.

Supported methods: `print.rcrologit`.

rcrologit *Estimation and Inference for Random Coefficients Rank-Ordered Logit*

## Description

This function takes the object prepared by dataPrep and estimates a random coefficient rank-ordered logit. The rank-ordered logit - sometimes termed *exploded logit model* - was originally proposed in Beggs et al. (1981) and it is an extension of the Luce (1959)-McFadden (1974) model.

These models automatically implies independence of irrelevant alternatives (Debreu 1960). However, McFadden and Train (2000) show that if agents are allowed to have heterogeneous tastes (i.e., random coefficients), then the conditional probability choices induced by the latent utility model can match those implied by virtually any discrete choice probability model.

The package rcrologit, depending on the type of covariates specified in dataPrep, allows the user to estimate:

- the standard rank-ordered logit model if either covs.fix or covsInt.fix are specified.
- the random coefficients rank-ordered logit model if either covs.het or covsInt.het are specified.
- the mixed random coefficients rank-ordered logit model if at least one of covs.fix or covsInt.fix and at least one of covs.het or covsInt.het are specified.

For more information on the underlying specification see the official repository.

## Usage

```
rcrologit(
  dataprep,
  Sigma = "diagonal",
  S = 50L,
  approx.method = "MC",
  bias.correction = FALSE,
  robust = FALSE,
  stdErr.dfadj = TRUE,
  skip.stdErr = FALSE,
  Ncores = 1L,
  verbose = TRUE,
  control.opts = NULL
)
```

## Arguments

| | |
|---|---|
| dataprep | object of class 'rcrologit' prepared via dataPrep. |
| Sigma | structure of the variance-covariance of the random coefficients. It must be one of "diagonal" or "cholesky". Default is Sigma="diagonal". See **Details** section for more. |
| S | integer denoting the number of simulations when approximating the integrals in the conditional choice probabilities. Default is S=50. More on how to select the optimal $S$ can be found in (Cameron and Trivedi 2005) and (Hajivassiliou 2000). |

`approx.method` string indicating the procedure to approximate the integrals in the conditional choice probabilities when including random coefficients. At the moment only approximation via monte-carlo simulation is available. Future releases will include importance sampling and other alternatives

`bias.correction` if `TRUE` applies the bias correction for simulated maximum likelihood proposed in (Gouriéroux and Monfort 1991). For more details see Section 12.4.4 in (Cameron and Trivedi 2005). This option is effective only when random coefficients are included in the model.

`robust` if `TRUE` computes standard errors robust to misspecification.

`stdErr.dfadj` boolean indicating whether a degrees-of-freedom correction should be used when estimating the variance of the ML estimator. See **Details** section for more.

`skip.stdErr` if `TRUE` skips computation of standard errors which can be intensive in the random coefficient rologit model.

`Ncores` integer indicating the number of cores to be used in simulating conditional choice probabilities. It affects speed only when random coefficients are included in the model. Speed gains are sensible whenever $S \geq 100$. On Windows systems it is set automatically to `Ncores = 1`.

`verbose` if `TRUE` prints additional information in the console.

`control.opts` a list containing options to be passed to the underlying optimizer `optim`.

## Details

- **Variance-Covariance degrees-of-freedom adjustment.** When `stdErr.dfadj = FALSE`, the estimate for the asymptotic variance of $\sqrt{N}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is divided by the sample size $N$. If instead, `stdErr.dfadj = TRUE`, then the estimate for the asymptotic variance of $\sqrt{N}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is divided by $N-k$, where $k$ is the dimension of $\boldsymbol{\theta}$.
- **Variance-Covariance of random parameters.** The option `Sigma` allows the user to model directly the covariance structure of the random coefficients. Precisely, it shape $\boldsymbol{\Sigma}$ in
$$\left[\begin{array}{c}\boldsymbol{\alpha}_i\\ \boldsymbol{\beta}_i\end{array}\right] \sim \mathsf{N}\left(\left[\begin{array}{c}\boldsymbol{\alpha}_{\mathtt{R}}\\ \boldsymbol{\beta}_{\mathtt{R}}\end{array}\right], \boldsymbol{\Sigma}\right).$$
In practice, to avoid issues with the positive-definiteness of $\boldsymbol{\Sigma}$ during the optimization routine, it is more convenient to model the random coefficients as
$$\left[\begin{array}{c}\boldsymbol{\alpha}_i\\ \boldsymbol{\beta}_i\end{array}\right] = \left[\begin{array}{c}\boldsymbol{\alpha}_{\mathtt{R}}\\ \boldsymbol{\beta}_{\mathtt{R}}\end{array}\right] + \Lambda\mathbf{u}_i, \quad \mathbf{u}_i \sim \mathsf{N}(0, \boldsymbol{I}).$$
The option `Sigma` directly models the shape of $\Lambda$ and allows the user to choose between a diagonal structure (`Sigma="diagonal"`) and a lower-triangular structure (`Sigma="cholesky"`).
- **Model.** The function estimates a random coefficient rank-ordered logit model induced by the latent utility model McFadden (1974)
$$U_{i\ell} = u_{i\ell} + \epsilon_{i\ell}, \quad i = 1, 2, \dots, n, \quad j = 0, 1, \dots, J.$$
In its most general form, we model $u_{i\ell}$ as
$$u_{i\ell} = X_{i\ell}^\top \boldsymbol{\beta}_{\mathtt{F}} + Z_i^\top \boldsymbol{\alpha}_{\ell,\mathtt{F}} + W_{i\ell}^\top \boldsymbol{\beta}_i + V_i^\top \boldsymbol{\alpha}_{i\ell} + \delta_\ell$$
where

- $X_{i\ell}$ are covariates varying at the unit-alternative level whose coefficients are modelled as fixed. The user can specify these covariates via the option `covs.fix`.
- $Z_i$ are covariates varying at the unit level whose coefficients are modelled as fixed. The user can specify these covariates via the option `covsInt.fix`. These coefficients are interacted with a dummy for the choice and treated as alternative varying covariates, i.e., $Z_{i\ell} = \sum_{j=1}^{J} Z_i \times \mathbf{1}(j = \ell)$, where $J=0$ is the reference group.
- $W_{i\ell}$ are covariates varying at the unit-alternative level whose coefficients are modelled as random. The user can specify these covariates via the option `covs.het`.
- $V_i$ are covariates varying at the unit level whose coefficients are modelled as random. The user can specify these covariates via the option `covsInt.het`. These coefficients are interacted with a dummy for the choice and treated as alternative varying covariates, i.e., $V_{i\ell} = \sum_{j=1}^{J} V_i \times \mathbf{1}(j = \ell)$, where $J=0$ is the reference group.
- the random coefficients are modeled as a joint multivariate normal and are i.i.d. across units,

$$\left[\begin{array}{c}\boldsymbol{\alpha}_i\\ \boldsymbol{\beta}_i\end{array}\right] \sim \mathsf{N}\left(\left[\begin{array}{c}\boldsymbol{\alpha}_{\mathtt{R}}\\ \boldsymbol{\beta}_{\mathtt{R}}\end{array}\right], \boldsymbol{\Sigma}\right)$$

- $\delta_\ell$ are alternative-specific fixed effects that can be specified via the option `FE`.
- $\epsilon_{i\ell} \sim \mathsf{Gu}(0,1)$are idiosyncratic i.i.d. shocks.

The parameter vector to be estimated is thus

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

where the first alternative-fixed effect has been normalized to 0.

For more information on the underlying specification see the official repository.

## Value

The function returns a list containing the following objects:

| | |
|---|---|
| `b` | vector containing all the estimated parameters |
| `bfix` | vector containing the estimated parameters corresponding to the "fixed" coefficients |
| `bhet` | vector containing the estimated parameters corresponding to the "random" coefficients |
| `Lambda` | loading matrix of the shocks to the random coefficients |
| `Sigma` | variance-covariance matrix of the estimated parameters |
| `param.spec` | a list containing some parameters describing the specification chosen by the user |
| `X` | a list containing the design matrix for each alternative |
| `fitted` | a dataframe containing the fitted values of the model. The fitted values are computed as the probability that alternatives are ranked first. |

## Author(s)

Chiara Motta, University of California Berkeley. <cmotta@berkeley.edu>

Filippo Palomba, Princeton University. <fpalomba@princeton.edu>

### References


Beggs S, Cardell S, Hausman J (1981). "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics*, **17**(1), 1–19. ISSN 0304-4076, doi:10.1016/03044076(81)900567, 2023-03-09.

Cameron AC, Trivedi PK (2005). *Microeconometrics: Methods and Applications*. Cambridge university press.

Debreu G (1960). "Review of RD Luce, Individual Choice Behavior: A Theoretical Analysis." *American Economic Review*, **50**(1), 186–188.

Gouriéroux C, Monfort A (1991). "Simulation Based Inference in Models with Heterogeneity." *Annales d'Économie et de Statistique*, 69–107. ISSN 0769-489X, doi:10.2307/20075807, 20075807, 2023-05-28.

Hajivassiliou VA (2000). *Some Practical Issues in Maximum Simulated Likelihood*. Suntory-Toyota International Centre for Economics and Related Disciplines . . . .

Luce RD (1959). *Individual Choice Behavior: A Theoretical Analysis*. Courier Corporation. ISBN 978-0-486-15339-1.

McFadden D (1974). "Frontiers in Econometrics." In *Conditional Logit Analysis of Qualitative Choice Behavior*, 105–142. Academic Press. ISBN 0-12-776150-0.

McFadden D, Train K (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, **15**(5), 447–470. ISSN 1099-1255, doi:10.1002/10991255(200009/10)15:5<447::AID-JAE570>3.0.CO;21, 2023-05-04.


### See Also

dataPrep

### Examples

```
data <- rcrologit_data
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)
```

---

## rcrologit_data — *Simulated Data According to McFadden (1974) Latent Utility Model*

---

### Description

A dataset containing three-option rankings of 3000 units and some covariates

### Usage

```
data(rcrologit_data)
```

### Format

A data frame with 9000 rows and 11 variables:

**Worker_ID** worker id.

**alternative** job offer id (within worker).

**rank** rank of the job offer.

**Firm_ID** id of the firm offering the job.

**Gender** gender of worker (1 if Female).

**Educ** education level of worker (1 if <HS, 2 if HS, 3 if College or more).

**WorkerSector** sector of worker's previous occupation (1 if Manufacturing, 0 otherwise).

**Wage** wage offered.

**log_Wage** log of wage offered.

**FirmSector** sector of the firm offering the job (1 if Manufacturing, 0 otherwise).

---

## `summary.rcrologit` *Summary Method for Random Coefficients Rank-Ordered Logit*

---

### Description

The summary method for for random coefficients rank-ordered logit fitted objects.

### Usage

```
## S3 method for class 'rcrologit'
summary(object, printFE = FALSE, ...)
```

### Arguments

| | |
|---|---|
| `object` | Class "rcrologit" object, obtained by calling `rcrologit`. |
| `printFE` | Whether fixed effects coefficients and standard errors should be displayed. |
| `...` | Other arguments. |

### Value

No return value, called to print `rcrologit` results.

### Author(s)

Chiara Motta, University of California Berkeley. <cmotta@berkeley.edu>

Filippo Palomba, Princeton University. <fpalomba@princeton.edu>

### See Also

`rcrologit` for estimation and inference for random coefficients rank-ordered logit.

Supported methods: `print.rcrologit`.

# Index