Package 'PanelCount'

October 12, 2022

Type Package

Title Random Effects and/or Sample Selection Models for Panel Count Data

Version 2.0.0

Date 2022-10-05

Author Jing Peng

Maintainer Jing Peng <jing.peng@uconn.edu>

Description A high performance package implementing random effects and/or sample selection models for panel count data. The details of the models are discussed in Peng and Van den Bulte (2022) <doi:10.2139/ssrn.2702053>.

License MIT + file LICENSE

Encoding UTF-8

VignetteBuilder knitr

LazyData TRUE

Depends R (>= 3.5.0)

Imports Rcpp, statmod, MASS

Suggests knitr, rmarkdown

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

RoxygenNote 7.2.1

Repository CRAN

Date/Publication 2022-10-07 06:10:02 UTC

R topics documented:

`PanelCount` *Panel Count Models with Random Effects and/or Sample Selection*

Description

A high performance package for estimating panel count models with random effects and/or sample selection.

Functions

ProbitRE: Probit model with random effects on individuals

PoissonRE: Poisson model with random effects on individuals

PLN_RE: Poisson Lognormal model with random effects on individuals

ProbitRE_PoissonRE: PoissonRE and ProbitRE model with correlated random effects on individuals

ProbitRE_PLNRE: PLN_RE and ProbitRE model with correlated random effects on individual level and correlated error terms on individual-time level

References

1. Peng, J., & Van den Bulte, C. (2022). Participation vs. Effectiveness in Sponsored Tweet Campaigns: A Quality-Quantity Conundrum. Available at SSRN: https://ssrn.com/abstract=2702053

2. Peng, J., & Van Den Bulte, C. (2015). How to Better Target and Incent Paid Endorsers in Social Advertising Campaigns: A Field Experiment. 2015 International Conference on Information Systems. https://aisel.aisnet.org/icis2015/proceedings/SocialMedia/24

PLN_RE *A Poisson Lognormal Model with Random Effects*

Description

Estimate a Poisson model with random effects at the individual and individual-time levels.

$$E[y_{it}|x_{it}, v_i, \epsilon_{it}] = exp(\boldsymbol{\beta}\mathbf{x_{it}}' + \sigma v_i + \gamma\epsilon_{it})$$

Notations:

- x_{it}: variables influencing the selection decision y_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- v_i: individual level random effect
- ϵ_{it}: individual-time level random effect

v_i and ϵ_{it} can both account for overdispersion.

Usage

```
PLN_RE(
  formula,
  data,
  id.name,
  par = NULL,
  sigma = NULL,
  gamma = NULL,
  method = "BFGS",
  adaptiveLL = TRUE,
  stopUpdate = FALSE,
  se_type = c("BHHH", "Hessian")[1],
  H = 12,
  psnH = 12,
  reltol = sqrt(.Machine$double.eps),
  verbose = 0
)
```

Arguments

formula	Formula of the model
data	Input data, a data.frame object
id.name	The name of the column representing id. Data will be sorted by id to improve estimation speed.
par	Starting values for estimates. Default to estimates of Poisson RE model.
sigma	Starting value for sigma. Defaults to 1 and will be ignored if par is provided.
gamma	Starting value for gamma. Defaults to 1 and will be ignored if par is provided.

`method`	Optimization method used by optim. Defaults to 'BFGS'.
`adaptiveLL`	Whether to use Adaptive Gaussian Quadrature. Defaults to TRUE because it is more reliable (though slower) for long panels.
`stopUpdate`	Whether to disable update of Adaptive Gaussian Quadrature parameters. Defaults to FALSE.
`se_type`	Report Hessian or BHHH standard errors. Defaults to BHHH.
`H`	Number of Quadrature points used for numerical integration using the Gaussian-Hermite Quadrature method. Defaults to 20.
`psnH`	Number of Quadrature points for Poisson RE model
`reltol`	Relative convergence tolerance. The algorithm stops if it is unable to reduce the value by a factor of reltol * (abs(val) + reltol) at a step. Defaults to sqrt(.Machine$double.eps), typically about 1e-8.
`verbose`	A integer indicating how much output to display during the estimation process. • <0 - No ouput • 0 - Basic output (model estimates) • 1 - Moderate output, basic ouput + parameter and likelihood in each iteration • 2 - Extensive output, moderate output + gradient values on each call

Value

A list containing the results of the estimated model, some of which are inherited from the return of optim

- estimates: Model estimates with 95% confidence intervals
- par: Point estimates
- var_bhhh: BHHH covariance matrix, inverse of the outer product of gradient at the maximum
- var_hessian: Inverse of negative Hessian matrix (the second order derivative of likelihood at the maximum)
- se_bhhh: BHHH standard errors
- g: Gradient function at maximum
- gtHg: $g'H^-1g$, where H^-1 is approximated by var_bhhh. A value close to zero (e.g., <1e-3 or 1e-6) indicates good convergence.
- LL: Likelihood
- AIC: AIC
- BIC: BIC
- n_obs: Number of observations
- time: Time takes to estimate the model
- partial: Average partial effect at the population level
- paritalAvgObs: Partial effect for an individual with average characteristics
- predict: A list with predicted participation probability (prob), predicted potential outcome (outcome), and predicted actual outcome (actual_outcome).

- counts: From optim. A two-element integer vector giving the number of calls to fn and gr respectively. This excludes those calls needed to compute the Hessian, if requested, and any calls to fn to compute a finite-difference approximation to the gradient.
- message: From optim. A character string giving any additional information returned by the optimizer, or NULL.
- convergence: From optim. An integer code. 0 indicates successful completion. Note that the list inherits all the complements in the output of optim. See the documentation of optim for more details.

References

1. Peng, J., & Van den Bulte, C. (2022). Participation vs. Effectiveness in Sponsored Tweet Campaigns: A Quality-Quantity Conundrum. Available at SSRN: https://ssrn.com/abstract=2702053
2. Peng, J., & Van Den Bulte, C. (2015). How to Better Target and Incent Paid Endorsers in Social Advertising Campaigns: A Field Experiment. 2015 International Conference on Information Systems. https://aisel.aisnet.org/icis2015/proceedings/SocialMedia/24

See Also

Other PanelCount: PoissonRE(), ProbitRE_PLNRE(), ProbitRE_PoissonRE(), ProbitRE()

Examples

```
# Use the simulated dataset, in which the true coefficient of x is 1.
# Estimated coefficient is biased due to omission of self-selection
data(sim)
res = PLN_RE(y~x, data=sim[!is.na(sim$y), ], id.name='id', verbose=-1)
res$estimates
```

PoissonRE *A Poisson Model with Random Effects*

Description

Estimate a Poisson model with random effects at the individual level.

$$E[y_{it}|x_{it}, v_i] = exp(\boldsymbol{\beta}\mathbf{x_{it}}' + \sigma v_i)$$

Notations:

- x_{it}: variables influencing the outcome y_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- v_i: individual level random effect

Usage

```
PoissonRE(
  formula,
  data,
  id.name,
  par = NULL,
  sigma = NULL,
  method = "BFGS",
  stopUpdate = FALSE,
  se_type = c("Hessian", "BHHH")[1],
  H = 20,
  reltol = sqrt(.Machine$double.eps),
  verbose = 0
)
```

Arguments

formula	Formula of the model
data	Input data, a data.frame object
id.name	The name of the column representing id. Data will be sorted by id to improve estimation speed.
par	Starting values for estimates. Default to estimates of Poisson Model
sigma	Starting value for sigma. Defaults to 1 and will be ignored if par is provided.
method	Optimization method used by optim. Defaults to 'BFGS'.
stopUpdate	Whether to disable update of Adaptive Gaussian Quadrature parameters. Defaults to FALSE.
se_type	Report Hessian or BHHH standard errors. Defaults to Hessian.
H	Number of Quadrature points used for numerical integration using the Gaussian-Hermite Quadrature method. Defaults to 20.
reltol	Relative convergence tolerance. The algorithm stops if it is unable to reduce the value by a factor of reltol * (abs(val) + reltol) at a step. Defaults to sqrt(.Machine$double.eps), typically about 1e-8.
verbose	A integer indicating how much output to display during the estimation process. • <0 - No ouput • 0 - Basic output (model estimates) • 1 - Moderate output, basic ouput + parameter and likelihood in each iteration • 2 - Extensive output, moderate output + gradient values on each call

Value

A list containing the results of the estimated model, some of which are inherited from the return of optim

- estimates: Model estimates with 95% confidence intervals

- par: Point estimates
- var_bhhh: BHHH covariance matrix, inverse of the outer product of gradient at the maximum
- var_hessian: Inverse of negative Hessian matrix (the second order derivative of likelihood at the maximum)
- se_bhhh: BHHH standard errors
- g: Gradient function at maximum
- gtHg: $g'H^-1g$, where H^-1 is approximated by var_bhhh. A value close to zero (e.g., <1e-3 or 1e-6) indicates good convergence.
- LL: Likelihood
- AIC: AIC
- BIC: BIC
- n_obs: Number of observations
- time: Time takes to estimate the model
- partial: Average partial effect at the population level
- paritalAvgObs: Partial effect for an individual with average characteristics
- predict: A list with predicted participation probability (prob), predicted potential outcome (outcome), and predicted actual outcome (actual_outcome).
- counts: From optim. A two-element integer vector giving the number of calls to fn and gr respectively. This excludes those calls needed to compute the Hessian, if requested, and any calls to fn to compute a finite-difference approximation to the gradient.
- message: From optim. A character string giving any additional information returned by the optimizer, or NULL.
- convergence: From optim. An integer code. 0 indicates successful completion. Note that the list inherits all the complements in the output of optim. See the documentation of optim for more details.

References

1. Peng, J., & Van den Bulte, C. (2022). Participation vs. Effectiveness in Sponsored Tweet Campaigns: A Quality-Quantity Conundrum. Available at SSRN: https://ssrn.com/abstract=2702053
2. Peng, J., & Van Den Bulte, C. (2015). How to Better Target and Incent Paid Endorsers in Social Advertising Campaigns: A Field Experiment. 2015 International Conference on Information Systems. https://aisel.aisnet.org/icis2015/proceedings/SocialMedia/24

See Also

Other PanelCount: `PLN_RE()`, `ProbitRE_PLNRE()`, `ProbitRE_PoissonRE()`, `ProbitRE()`

Examples

```
# Use the simulated dataset, in which the true coefficient of x is 1.
# Estimated coefficient is biased primarily due to omission of self-selection
data(sim)
res = PoissonRE(y~x, data=sim[!is.na(sim$y), ], id.name='id', verbose=-1)
res$estimates
```

predict_ProbitRE_PLNRE *Predictions of CRE_SS model on new sample*

Description

Predictions of CRE_SS model on new sample. Please make sure the factor variables in the test data do not have levels not shown in the training data.

Usage

```
predict_ProbitRE_PLNRE(
  par,
  sel_form,
  out_form,
  data,
  offset_w_name = NULL,
  offset_x_name = NULL
)
```

Arguments

par	Model estimates
sel_form	Formula for selection equation, a Probit model with random effects
out_form	Formula for outcome equation, a Poisson Lognormal model with random effects
data	Input data, a data.frame object
offset_w_name	Offset variables in selection equation, if any.
offset_x_name	Offset variables in outcome equation, if any.

Value

A list with three sets of predictions

- prob: Predicted probability to participate
- outcome: Predicted potential outcome
- actual_outcome: Predicted actual outcome

predict_ProbitRE_PoissonRE *Predictions of CRE model on new sample*

Description

Predictions of CRE model on new sample. Please make sure the factor variables in the test data do not have levels not shown in the training data.

Usage

```
predict_ProbitRE_PoissonRE(
  par,
  sel_form,
  out_form,
  data,
  offset_w_name = NULL,
  offset_x_name = NULL
)
```

Arguments

par	Model estimates
sel_form	Formula for selection equation, a Probit model with random effects
out_form	Formula for outcome equation, a Poisson Lognormal model with random effects
data	Input data, a data.frame object
offset_w_name	Offset variables in selection equation, if any.
offset_x_name	Offset variables in outcome equation, if any.

Value

A list with three sets of predictions

- prob: Predicted probability to participate
- outcome: Predicted potential outcome
- actual_outcome: Predicted actual outcome

ProbitRE *A Probit Model with Random Effects*

Description

Estimate a Probit model with random effects at the individual level.

$$z_{it} = 1(\boldsymbol{\alpha}\mathbf{w_{it}}' + \delta u_i + \xi_{it} > 0)$$

Notations:

- w_{it}: variables influencing the selection decision z_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- u_i: individual level random effect
- ξ_{it}: error term

Usage

```
ProbitRE(
  formula,
  data,
  id.name,
  par = NULL,
  delta = NULL,
  method = "BFGS",
  se_type = c("Hessian", "BHHH")[1],
  H = 20,
  reltol = sqrt(.Machine$double.eps),
  verbose = 0
)
```

Arguments

formula	Formula of the model
data	Input data, a data.frame object
id.name	The name of the column representing id. Data will be sorted by id to improve estimation speed.
par	Starting values for estimates. Default to estimates of Probit model.
delta	Starting value for delta. Defaults to 1 and will be ignored if par is provided.
method	Optimization method used by optim. Defaults to 'BFGS'.
se_type	Report Hessian or BHHH standard errors. Defaults to Hessian.
H	Number of Quadrature points used for numerical integration using the Gaussian-Hermite Quadrature method. Defaults to 20.

`reltol` Relative convergence tolerance. The algorithm stops if it is unable to reduce the value by a factor of reltol * (abs(val) + reltol) at a step. Defaults to sqrt(.Machine$double.eps), typically about 1e-8.

`verbose` A integer indicating how much output to display during the estimation process.

- <0 - No ouput
- 0 - Basic output (model estimates)
- 1 - Moderate output, basic ouput + parameter and likelihood in each iteration
- 2 - Extensive output, moderate output + gradient values on each call

Value

A list containing the results of the estimated model, some of which are inherited from the return of optim

- estimates: Model estimates with 95% confidence intervals
- par: Point estimates
- var_bhhh: BHHH covariance matrix, inverse of the outer product of gradient at the maximum
- var_hessian: Inverse of negative Hessian matrix (the second order derivative of likelihood at the maximum)
- se_bhhh: BHHH standard errors
- g: Gradient function at maximum
- gtHg: $g'H^-1g$, where H^-1 is approximated by var_bhhh. A value close to zero (e.g., <1e-3 or 1e-6) indicates good convergence.
- LL: Likelihood
- AIC: AIC
- BIC: BIC
- n_obs: Number of observations
- time: Time takes to estimate the model
- partial: Average partial effect at the population level
- paritalAvgObs: Partial effect for an individual with average characteristics
- predict: A list with predicted participation probability (prob), predicted potential outcome (outcome), and predicted actual outcome (actual_outcome).
- counts: From optim. A two-element integer vector giving the number of calls to fn and gr respectively. This excludes those calls needed to compute the Hessian, if requested, and any calls to fn to compute a finite-difference approximation to the gradient.
- message: From optim. A character string giving any additional information returned by the optimizer, or NULL.
- convergence: From optim. An integer code. 0 indicates successful completion. Note that the list inherits all the complements in the output of optim. See the documentation of optim for more details.
- estimates model estimates with 95% confidence intervals

- par point estimates
- var_bhhh BHHH covariance matrix, inverse of the outer product of gradient at the maximum
- var_hessian Inverse of negative Hessian matrix (the second order derivative of likelihood at the maximum)
- se_bhhh BHHH standard errors
- g graident function at maximum
- LL likelihood
- AIC AIC
- BIC BIC
- n_obs Number of observations
- counts A two-element integer vector giving the number of calls to fn and gr respectively. This excludes those calls needed to compute the Hessian, if requested, and any calls to fn to compute a finite-difference approximation to the gradient.
- time Time takes to estimate the model
- message A character string giving any additional information returned by the optimizer, or NULL.
- convergence An integer code. 0 indicates successful completion. Note that the list inherits all the complements in the output of optim. See the documentation of optim for more details.

References

1. Peng, J., & Van den Bulte, C. (2022). Participation vs. Effectiveness in Sponsored Tweet Campaigns: A Quality-Quantity Conundrum. Available at SSRN: https://ssrn.com/abstract=2702053
2. Peng, J., & Van Den Bulte, C. (2015). How to Better Target and Incent Paid Endorsers in Social Advertising Campaigns: A Field Experiment. 2015 International Conference on Information Systems. https://aisel.aisnet.org/icis2015/proceedings/SocialMedia/24

See Also

Other PanelCount: `PLN_RE()`, `PoissonRE()`, `ProbitRE_PLNRE()`, `ProbitRE_PoissonRE()`

Examples

```
# Use the simulated dataset, in which the true coefficients of x and w are 1.
data(sim)
res = ProbitRE(z~x+w, data=sim, id.name='id', verbose=-1)
res$estimates
```

ProbitRE_PLNRE *Poisson Lognormal Model with Random Effects and Sample Selection*

Description

Estimates the following two-stage model:

Selection equation (ProbitRE - Probit model with individual level random effects):

$$z_{it} = 1(\boldsymbol{\alpha}\mathbf{w_{it}}' + \delta u_i + \xi_{it} > 0)$$

Outcome Equation (PLN_RE - Poisson Lognormal model with individual-time level random effects):

$$E[y_{it}|x_{it}, v_i, \epsilon_{it}] = exp(\boldsymbol{\beta}\mathbf{x_{it}}' + \sigma v_i + \gamma\epsilon_{it})$$

Correlation (self-selection at both individual and individual-time level):

- u_i and v_i are bivariate normally distributed with a correlation of ρ.
- ξ_{it} and ϵ_{it} are bivariate normally distributed with a correlation of τ.

Notations:

- w_{it}: variables influencing the selection decision z_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- x_{it}: variables influencing the outcome y_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- u_i: individual level random effect in the selection equation
- v_i: individual level random effect in the outcome equation
- ξ_{it}: error term in the selection equation
- ϵ_{it}: individual-time level random effect in the outcome equation

Usage

```
ProbitRE_PLNRE(
  sel_form,
  out_form,
  data,
  id.name,
  testData = NULL,
  par = NULL,
  disable_rho = FALSE,
  disable_tau = FALSE,
  delta = NULL,
  sigma = NULL,
  gamma = NULL,
  rho = NULL,
  tau = NULL,
```

```
  method = "BFGS",
  se_type = c("BHHH", "Hessian")[1],
  H = c(10, 10),
  psnH = 20,
  prbH = 20,
  plnreH = 20,
  reltol = sqrt(.Machine$double.eps),
  factr = 1e+07,
  verbose = 1,
  offset_w_name = NULL,
  offset_x_name = NULL
)
```

Arguments

sel_form	Formula for selection equation, a Probit model with random effects
out_form	Formula for outcome equation, a Poisson Lognormal model with random effects
data	Input data, a data.frame object
id.name	The name of the column representing id. Data will be sorted by id to improve estimation speed.
testData	Test data for prediction, a data.frame object
par	Starting values for estimates. Default to estimates of standalone selection and outcome models.
disable_rho	Whether to disable correlation at the individual level random effect. Defaults to FALSE.
disable_tau	Whether to disable correlation at the individual-time level random effect / error term. Defaults to FALSE.
delta	Starting value for delta. Will be ignored if par is provided.
sigma	Starting value for sigma. Will be ignored if par is provided.
gamma	Starting value for gamma. Will be ignored if par is provided.
rho	Starting value for rho. Defaults to 0 and will be ignored if par is provided.
tau	Starting value for tau. Defaults to 0 and will be ignored if par is provided.
method	Optimization method used by optim. Defaults to 'BFGS'.
se_type	Report Hessian or BHHH standard errors. Defaults to BHHH. Hessian matrix is extremely time-consuming to calculate numerically for large datasets.
H	A integer vector of length 2, specifying the number of points for inner and outer Quadratures
psnH	Number of Quadrature points for Poisson RE model
prbH	Number of Quadrature points for Probit RE model
plnreH	Number of Quadrature points for PLN_RE model
reltol	Relative convergence tolerance. The algorithm stops if it is unable to reduce the value by a factor of reltol * (abs(val) + reltol) at a step. Defaults to sqrt(.Machine$double.eps), typically about 1e-8.

`factr`	L-BFGS-B method uses factr instead of reltol to control for precision. Default is 1e7, that is a tolerance of about 1e-8.
`verbose`	A integer indicating how much output to display during the estimation process. • <0 - No ouput • 0 - Basic output (model estimates) • 1 - Moderate output, basic ouput + parameter and likelihood in each iteration • 2 - Extensive output, moderate output + gradient values on each call
`offset_w_name`	An offset variable whose coefficient is assumed to be 1 in the selection equation
`offset_x_name`	An offset variable whose coefficient is assumed to be 1 in the outcome equation

Value

A list containing the results of the estimated model, some of which are inherited from the return of optim

- estimates: Model estimates with 95% confidence intervals
- par: Point estimates
- var_bhhh: BHHH covariance matrix, inverse of the outer product of gradient at the maximum
- se_bhhh: BHHH standard errors
- g: Gradient function at maximum
- gtHg: $g'H^{-}1g$, where H^-1 is approximated by var_bhhh. A value close to zero (e.g., <1e-3 or 1e-6) indicates good convergence.
- LL: Likelihood
- AIC: AIC
- BIC: BIC
- n_obs: Number of observations
- time: Time takes to estimate the model
- partial: Average partial effect at the population level
- paritalAvgObs: Partial effect for an individual with average characteristics
- predict: A list with predicted participation probability (prob), predicted potential outcome (outcome), and predicted actual outcome (actual_outcome).
- counts: From optim. A two-element integer vector giving the number of calls to fn and gr respectively. This excludes those calls needed to compute the Hessian, if requested, and any calls to fn to compute a finite-difference approximation to the gradient.
- message: From optim. A character string giving any additional information returned by the optimizer, or NULL.
- convergence: From optim. An integer code. 0 indicates successful completion. Note that the list inherits all the complements in the output of optim. See the documentation of optim for more details.

References

1. Peng, J., & Van den Bulte, C. (2022). Participation vs. Effectiveness in Sponsored Tweet Campaigns: A Quality-Quantity Conundrum. Available at SSRN: https://ssrn.com/abstract=2702053
2. Peng, J., & Van Den Bulte, C. (2015). How to Better Target and Incent Paid Endorsers in Social Advertising Campaigns: A Field Experiment. 2015 International Conference on Information Systems. https://aisel.aisnet.org/icis2015/proceedings/SocialMedia/24

See Also

Other PanelCount: PLN_RE(), PoissonRE(), ProbitRE_PoissonRE(), ProbitRE()

Examples

```
# Use the simulated dataset, in which the true coefficients of x and w are 1 in both stages.
# The model can recover the true parameters very well
data(sim)
res = ProbitRE_PLNRE(z~x+w, y~x, data=sim, id.name='id')
res$estimates
```

`ProbitRE_PoissonRE` *Poisson RE model with Sample Selection*

Description

Estimates the following two-stage model

Selection equation (ProbitRE - Probit model with individual level random effects):

$$z_{it} = 1(\boldsymbol{\alpha}\mathbf{w_{it}}' + \delta u_i + \xi_{it} > 0)$$

Outcome Equation (PoissonRE - Poisson with individual level random effects):

$$E[y_{it}|x_{it}, v_i] = exp(\boldsymbol{\beta}\mathbf{x_{it}}' + \sigma v_i)$$

Correlation (self-selection at individual level):

- u_i and v_i are bivariate normally distributed with a correlation of ρ.

Notations:

- w_{it}: variables influencing the selection decision z_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- x_{it}: variables influencing the outcome y_{it}, which could be a mixture of time-variant variables, time-invariant variables, and time dummies
- u_i: individual level random effect in the selection equation
- v_i: individual level random effect in the outcome equation
- ξ_{it}: error term in the selection equation

Usage

```
ProbitRE_PoissonRE(
  sel_form,
  out_form,
  data,
  id.name,
  testData = NULL,
  par = NULL,
  delta = NULL,
  sigma = NULL,
  rho = NULL,
  method = "BFGS",
  se_type = c("BHHH", "Hessian")[1],
  H = c(10, 10),
  psnH = 20,
  prbH = 20,
  reltol = sqrt(.Machine$double.eps),
  verbose = 1,
  offset_w_name = NULL,
  offset_x_name = NULL
)
```

Arguments

sel_form	Formula for selection equation, a Probit model with random effects
out_form	Formula for outcome equation, a Poisson model with random effects
data	Input data, a data.frame object
id.name	The name of the column representing id. Data will be sorted by id to improve estimation speed.
testData	Test data for prediction, a data.frame object
par	Starting values for estimates. Default to estimates of standalone selection and outcome models.
delta	Starting value for delta. Will be ignored if par is provided.
sigma	Starting value for sigma. Will be ignored if par is provided.
rho	Starting value for rho. Defaults to 0 and will be ignored if par is provided.
method	Optimization method used by optim. Defaults to 'BFGS'.
se_type	Report Hessian or BHHH standard errors. Defaults to BHHH.
H	A integer vector of length 2, specifying the number of points for inner and outer Quadratures
psnH	Number of Quadrature points for Poisson RE model
prbH	Number of Quddrature points for Probit RE model
reltol	Relative convergence tolerance. The algorithm stops if it is unable to reduce the value by a factor of reltol * (abs(val) + reltol) at a step. Defaults to sqrt(.Machine$double.eps), typically about 1e-8.

verbose	A integer indicating how much output to display during the estimation process. • <0 - No ouput • 0 - Basic output (model estimates) • 1 - Moderate output, basic ouput + parameter and likelihood in each iteration • 2 - Extensive output, moderate output + gradient values on each call
offset_w_name	An offset variable whose coefficient is assumed to be 1 in the selection equation
offset_x_name	An offset variable whose coefficient is assumed to be 1 in the outcome equation

Value

A list containing the results of the estimated model, some of which are inherited from the return of optim

- estimates: Model estimates with 95% confidence intervals
- par: Point estimates
- var_bhhh: BHHH covariance matrix, inverse of the outer product of gradient at the maximum
- se_bhhh: BHHH standard errors
- g: Gradient function at maximum
- gtHg: $g'H^-1g$, where H^-1 is approximated by var_bhhh. A value close to zero (e.g., <1e-3 or 1e-6) indicates good convergence.
- LL: Likelihood
- AIC: AIC
- BIC: BIC
- n_obs: Number of observations
- time: Time takes to estimate the model
- partial: Average partial effect at the population level
- paritalAvgObs: Partial effect for an individual with average characteristics
- predict: A list with predicted participation probability (prob), predicted potential outcome (outcome), and predicted actual outcome (actual_outcome).
- counts: From optim. A two-element integer vector giving the number of calls to fn and gr respectively. This excludes those calls needed to compute the Hessian, if requested, and any calls to fn to compute a finite-difference approximation to the gradient.
- message: From optim. A character string giving any additional information returned by the optimizer, or NULL.
- convergence: From optim. An integer code. 0 indicates successful completion. Note that the list inherits all the complements in the output of optim. See the documentation of optim for more details.

References

1. Peng, J., & Van den Bulte, C. (2022). Participation vs. Effectiveness in Sponsored Tweet Campaigns: A Quality-Quantity Conundrum. Available at SSRN: https://ssrn.com/abstract=2702053
2. Peng, J., & Van Den Bulte, C. (2015). How to Better Target and Incent Paid Endorsers in Social Advertising Campaigns: A Field Experiment. 2015 International Conference on Information Systems. https://aisel.aisnet.org/icis2015/proceedings/SocialMedia/24

See Also

Other PanelCount: PLN_RE(), PoissonRE(), ProbitRE_PLNRE(), ProbitRE()

Examples

```
# Use the simulated dataset, in which the true coefficients of x and w are 1 in both stages.
# The simulated dataset includes self-selection at both individual and individual-time level,
# but this model only considers self-selection at the individual level.
data(sim)
res = ProbitRE_PoissonRE(z~x+w, y~x, data=sim, id.name='id')
res$estimates
```

---

sim *Simulated dataset with self-selection at both individual and individual-time level*

Description

A simulated dataset with 200 individuals and 10 periods. The true data generating process is the following:

Selection equation (ProbitRE - Probit model with individual level random effects):

$$z_{it} = 1(1 + x_{it} + w_{it} + u_i + \xi_{it} > 0$$

Outcome Equation (PLN_RE - Poisson Lognormal model with individual-time level random effects):

$$E[y_{it}|x_{it}, v_i, \epsilon_{it}] = exp(-1 + x_{it} + v_i + \epsilon_{it})$$

Correlation (self-selection at both individual and individual-time level):

- u_i and v_i are bivariate normally distributed with a correlation of 0.25.
- ξ_{it} and ϵ_{it} are bivariate normally distributed with a correlation of 0.5.

Usage

```
sim
```

Format

A simulated dataset with 200 individuals and 10 periods.

id id, from 1-200

time Time periods, from 1-10

z Whether an individual is selected in a given period. Outcome is observed only when z=1

y The outcome of an individual in a given period

x A covariate influencing both z and y, with true effects being 1

w A covariate influencing only z, with true effect being 1

Index