Package 'RegCalReliab'

October 6, 2025

Type Package

Title Regression Calibration Using Reliability Studies

Version 0.2.0

Maintainer Bowen Liu <bowenliu@hsph.harvard.edu>

Description Implements regression calibration methods for correcting measurement error in regression models using external or internal reliability studies. Methods are described in Carroll, Ruppert, Stefanski, and Crainiceanu (2006) ``Measurement Error in Nonlinear Models: A Modern Perspective" <doi:10.1201/9781420010138>.

License MIT + file LICENSE

Encoding UTF-8

Imports stats, sandwich

Suggests mgcv, knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 7.3.2

URL https://lbw080526.github.io/RegCalReliab/, https://github.com/lbw080526/RegCalReliab

BugReports https://github.com/lbw080526/RegCalReliab/issues

NeedsCompilation no

Author Bowen Liu [aut, cre, cph], Yu Lu [aut], Molin Wang [aut]

Repository CRAN

Date/Publication 2025-10-06 08:00:19 UTC

Contents

RC_ExReliab 2
RC_InReliab 4

Index **7**

RC_ExReliab *Unified Regression Calibration Wrapper (External Reliability Study)*

Description

A single formula interface for regression calibration in external reliability studies. The user simply specifies ‘link = "linear"‘, ‘"logistic"‘, or ‘"log"‘, and the wrapper selects the appropriate model: * ‘"linear"‘ → Gaussian (identity link) * ‘"logistic"‘ → Binomial (logit link) * ‘"log"‘ → Poisson (log link)

Usage

```
RC_ExReliab(
  formula,
  main_data,
  rep_data,
  link = c("linear", "logistic", "log"),
  return_details = FALSE
)
```

Arguments

formula	A formula or character string such as ‘Y ~ sbp(sbp2, sbp3) + chol(chol2, chol3) + age + weight‘. Terms of the form ‘var(rep1, rep2, ...)‘ are treated as error-prone exposures with replicates in ‘rep_data‘; other terms are treated as covariates W.
main_data	Data frame holding the outcome, error-prone exposures, and covariates.
rep_data	Data frame holding replicate columns referenced in ‘formula‘.
link	Character; one of ‘"linear"‘, ‘"logistic"‘, or ‘"log"‘.
return_details	Logical; if ‘TRUE‘, return parsed, prepared, and RC internals.

Value

A list with: * ‘uncorrected‘: naive regression estimates * ‘corrected‘ : sandwich-corrected regression calibration estimates * optional ‘details‘ if ‘return_details = TRUE‘

Examples

```
library(mgcv)
set.seed(123)
add_err <- function(v, sd = sqrt(0.4)) v + rnorm(length(v), 0, sd)

## --- Example 1: External 1Z 0W ---
x <- rnorm(3000)
z.main <- x[1:1500] + rnorm(1500, 0, sqrt(0.4))
z_rep <- rbind(
  cbind(add_err(x[1501:2000]), add_err(x[1501:2000]), NA, NA),
  cbind(add_err(x[2001:2400]), add_err(x[2001:2400]), add_err(x[2001:2400]), NA),
```

```
  cbind(add_err(x[2401:3000]), add_err(x[2401:3000]),
        add_err(x[2401:3000]), add_err(x[2401:3000]))
)
colnames(z_rep) <- paste0("z_", 1:4)
Y <- rbinom(1500, 1, plogis(-2.3 + log(1.5) * x[1:1500]))
main_data <- data.frame(Y = Y, z = z.main)
rep_data  <- data.frame(z_rep, check.names = FALSE)
res1 <- RC_ExReliab(Y ~ z(z_1, z_2, z_3, z_4), main_data, rep_data, link = "logistic")
res1$corrected

## --- Example 2: External 1Z 1W ---
x <- rnorm(3000)
W_main <- rnorm(1500)
W_rep  <- rnorm(1500)
z.main <- x[1:1500] + rnorm(1500, 0, sqrt(0.4))
z_rep <- rbind(
  cbind(add_err(x[1501:2000]), add_err(x[1501:2000]), NA, NA),
  cbind(add_err(x[2001:2400]), add_err(x[2001:2400]), add_err(x[2001:2400]), NA),
  cbind(add_err(x[2401:3000]), add_err(x[2401:3000]),
        add_err(x[2401:3000]), add_err(x[2401:3000]))
)
colnames(z_rep) <- paste0("z_", 1:4)
Y <- rbinom(1500, 1, plogis(-2.3 + log(1.5) * x[1:1500] + 0.5 * W_main))
main_data <- data.frame(Y = Y, z = z.main, W = W_main)
rep_data  <- data.frame(z_rep, W = W_rep, check.names = FALSE)
res2 <- RC_ExReliab(Y ~ z(z_1, z_2, z_3, z_4) + W, main_data, rep_data, link = "logistic")
res2$corrected

## --- Example 3: External 2Z 0W ---
x <- mgcv::rmvn(3000, c(0, 0), matrix(c(1, 0.3, 0.3, 1), 2))
z.main <- x[1:1500, ] + matrix(rnorm(1500 * 2, 0, sqrt(0.4)), 1500, 2)
colnames(z.main) <- c("z1", "z2")
z1_rep <- rbind(
  cbind(add_err(x[1501:2000, 1]), add_err(x[1501:2000, 1]), NA, NA),
 cbind(add_err(x[2001:2400, 1]), add_err(x[2001:2400, 1]), add_err(x[2001:2400, 1]), NA),
  cbind(add_err(x[2401:3000, 1]), add_err(x[2401:3000, 1]),
        add_err(x[2401:3000, 1]), add_err(x[2401:3000, 1]))
)
colnames(z1_rep) <- paste0("z1_", 1:4)
z2_rep <- rbind(
  cbind(add_err(x[1501:2000, 2]), add_err(x[1501:2000, 2]), NA, NA),
 cbind(add_err(x[2001:2400, 2]), add_err(x[2001:2400, 2]), add_err(x[2001:2400, 2]), NA),
  cbind(add_err(x[2401:3000, 2]), add_err(x[2401:3000, 2]),
        add_err(x[2401:3000, 2]), add_err(x[2401:3000, 2]))
)
colnames(z2_rep) <- paste0("z2_", 1:4)
Y <- rbinom(1500, 1, plogis(-2.3 + log(1.5) * rowSums(x[1:1500, ])))
main_data <- data.frame(Y = Y, z1 = z.main[, 1], z2 = z.main[, 2])
rep_data  <- data.frame(z1_rep, z2_rep, check.names = FALSE)
res3 <- RC_ExReliab(
  Y ~ z1(z1_1, z1_2, z1_3, z1_4) + z2(z2_1, z2_2, z2_3, z2_4),
  main_data, rep_data, link = "logistic"
)
```

```
res3$corrected
```

RC_InReliab *Unified Regression Calibration Wrapper (Internal Reliability Study)*

Description

A single formula interface for regression calibration in internal reliability studies. The user simply specifies ‘link = "linear"‘, ‘"logistic"‘, or ‘"log"‘, and the wrapper selects the appropriate model: * ‘"linear"‘ → Gaussian (identity link) * ‘"logistic"‘ → Binomial (logit link) * ‘"log"‘ → Poisson (log link)

Usage

```
RC_InReliab(
  formula,
  main_data,
  link = c("linear", "logistic", "log"),
  return_details = FALSE
)
```

Arguments

formula	A formula or character string such as ‘Y ~ sbp(sbp2, sbp3) + chol(chol2, chol3) + age + weight‘. Terms of the form ‘var(rep1, rep2, ...)‘ are treated as error-prone exposures with replicates in ‘main_data‘; other terms are treated as covariates W.
main_data	Data frame holding the outcome, replicate error-prone exposures, and any covariates.
link	Character; one of ‘"linear"‘, ‘"logistic"‘, or ‘"log"‘.
return_details	Logical; if ‘TRUE‘, return parsed, prepared, and RC internals.

Value

A list with: * ‘uncorrected‘: naive regression estimates * ‘corrected‘ : sandwich-corrected regression calibration estimates * optional ‘details‘ if ‘return_details = TRUE‘

Examples

```
set.seed(123)
add_err <- function(v, sd = sqrt(0.4)) v + rnorm(length(v), 0, sd)

## --- Example 1: Internal 1Z 0W ---
x <- rnorm(3000)
z <- rbind(
  cbind(add_err(x[1:1500]), NA, NA, NA),
```

```
  cbind(add_err(x[1501:2000]), add_err(x[1501:2000]), NA, NA),
  cbind(add_err(x[2001:2400]), add_err(x[2001:2400]), add_err(x[2001:2400]), NA),
  cbind(add_err(x[2401:3000]), add_err(x[2401:3000]),
        add_err(x[2401:3000]), add_err(x[2401:3000]))
)
colnames(z) <- paste0("z_", 1:4)
Y <- rbinom(3000, 1, plogis(-2.65 + log(1.5) * x))
main_data <- data.frame(Y, z)
res1 <- RC_InReliab(Y ~ myz(z_1, z_2, z_3, z_4),
                    main_data = main_data,
                    link = "logistic")
res1$corrected

## --- Example 2: Internal 1Z 1W ---
x  <- rnorm(3000)
W1 <- rnorm(3000)
z <- rbind(
  cbind(add_err(x[1:1500]), NA, NA, NA),
  cbind(add_err(x[1501:2000]), add_err(x[1501:2000]), NA, NA),
  cbind(add_err(x[2001:2400]), add_err(x[2001:2400]), add_err(x[2001:2400]), NA),
  cbind(add_err(x[2401:3000]), add_err(x[2401:3000]),
        add_err(x[2401:3000]), add_err(x[2401:3000]))
)
colnames(z) <- paste0("z_", 1:4)
Y <- rbinom(3000, 1, plogis(-2.65 + log(1.5) * x + 0.5 * W1))
main_data <- data.frame(Y, z, W1)
res2 <- RC_InReliab(Y ~ myz(z_1, z_2, z_3, z_4) + W1,
                    main_data = main_data,
                    link = "logistic")
res2$corrected

## --- Example 3: Internal 2Z 0W ---
x <- mgcv::rmvn(3000, c(0,0), matrix(c(1,0.3,0.3,1), 2))
z1 <- rbind(
  cbind(add_err(x[1:1500, 1]), NA, NA, NA),
  cbind(add_err(x[1501:2000, 1]), add_err(x[1501:2000, 1]), NA, NA),
 cbind(add_err(x[2001:2400, 1]), add_err(x[2001:2400, 1]), add_err(x[2001:2400, 1]), NA),
  cbind(add_err(x[2401:3000, 1]), add_err(x[2401:3000, 1]),
        add_err(x[2401:3000, 1]), add_err(x[2401:3000, 1]))
)
colnames(z1) <- paste0("z1_", 1:4)
z2 <- rbind(
  cbind(add_err(x[1:1500, 2]), NA, NA, NA),
  cbind(add_err(x[1501:2000, 2]), add_err(x[1501:2000, 2]), NA, NA),
 cbind(add_err(x[2001:2400, 2]), add_err(x[2001:2400, 2]), add_err(x[2001:2400, 2]), NA),
  cbind(add_err(x[2401:3000, 2]), add_err(x[2401:3000, 2]),
        add_err(x[2401:3000, 2]), add_err(x[2401:3000, 2]))
)
colnames(z2) <- paste0("z2_", 1:4)
Y <- rbinom(3000, 1, plogis(-2.65 + log(1.5) * rowSums(x)))
main_data <- data.frame(Y, z1, z2)
res3 <- RC_InReliab(
  Y ~ myz1(z1_1, z1_2, z1_3, z1_4) + myz2(z2_1, z2_2, z2_3, z2_4),
```

```
  main_data = main_data,
  link = "logistic")
res3$corrected
```

Index

RC_ExReliab, 2
RC_InReliab, 4