

R-Package ‘MrImputation II’

January 1, 2026

Type Package

Version 0.0.2

Title Multiple Ratio Imputation II

Date January 1, 2026

Author Masayoshi Takahashi (mtakahashi615@g.chuo-u.ac.jp)

Depends R (>=4.5.0)

Description R-Package MrImputation II implements the methods proposed by Takahashi (2025) and Takahashi (2026). This is an extended version of Takahashi (2017a, 2017b), which originally implemented R-Package MrImputation.

URL <https://github.com/mtakahashi123>

Repository GitHub

R topics documented

MrImputation-package	2
BPTtest.....	2
ExampleData00.csv.....	3
ExampleData05.csv.....	4
ExampleData10.csv.....	4
mranalyze2	5
mrdiag.....	6
mrimpute2	6
outputimp.csv.....	8
thetahat	8

MrImputation- package	<i>Multiple generalized ratio imputation with diagnostic tests</i>
--------------------------	--

Description

To use this package, click “Code” and “Download ZIP” at <https://github.com/mtakahashi123>. After downloading the package, set the working directory in R, and read R-Package MrImputationII using R-function `source` by either of the following methods.

Method 1

```
setwd("C:/Folder")
source("MrImputationII.R")
```

Method 2

```
source(file.choose())
```

References

- Takahashi, M. (2026) “Multiple generalized ratio imputation for missing data in official economic statistics: A flexible ratio estimator that automatically specifies the degree of heteroskedasticity.” Under review.
- Takahashi, M. 2025. “The treatment of missing values in official statistics.” *The Journal of Economics* 65, no. 5/6: 125-136.
- Takahashi, M. (2017a) “Multiple ratio imputation by the EMB algorithm: Theory and simulation.” *Journal of Modern Applied Statistical Methods*, 16 (1), 630-656.
- Takahashi, M. (2017b) “JMASM44: Implementing multiple ratio imputation by the EMB Algorithm (R).” *Journal of Modern Applied Statistical Methods*, 16 (1), 657-673.

BPTtest	<i>R-function to diagnose the estimated theta</i>
---------	---

Description

BPTtest implements the Breusch-Pagan-Takahashi test for the degree of heteroskedasticity.

Usage

```
BPTtest(y, x, theta = 0)
```

Arguments

<code>y</code>	Target variable for imputation
<code>x</code>	Auxiliary variable for the imputation model
<code>theta</code>	Value for the null hypothesis, i.e., the degree of heteroskedasticity. Default is 0.

Value

<code>BPT</code>	LM statistic for the BPT test for heteroskedasticity
<code>df</code>	Degrees of freedom
<code>p.value</code>	P-value. If the significance level is 0.05 and the p-value is less than 0.05, we reject the null hypothesis at the 5% error level, meaning that the estimated theta is considered to be wrong.

References

Takahashi, M. (2026) “Multiple generalized ratio imputation for missing data in official economic statistics: A flexible ratio estimator that automatically specifies the degree of heteroskedasticity.” Under review.

Takahashi, M. 2025. “The treatment of missing values in official statistics.” *The Journal of Economics* 65, no. 5/6: 125-136.

Breusch, T. S. and Pagan, A. R. (1979) “A simple test for heteroscedasticity and random coefficient variation.” *Econometrica*, 47 (5), 1287-1294.

Example

```
BPTtest(var1, var2, theta = 1.8956)
```

ExampleData00.csv *Simulated data for example 1*

Description

This is a simulated dataset as an example. The true value of theta is set to 0.0.

Usage

```
data1 <- read.csv("ExampleData00.csv", header=TRUE)
attach(data1)
```

Format

A data frame with 1000 observations on the following two variables.

var1 the target variable for imputation with 285 observations missing.

var2 the auxiliary variable for the imputation model.

ExampleData05.csv *Simulated data for example 2*

Description

This is a simulated dataset as an example. The true value of theta is set to 0.5.

Usage

```
data1 <- read.csv("ExampleData05.csv", header=TRUE)
attach(data1)
```

Format

A data frame with 1000 observations on the following two variables.

var1 the target variable for imputation with 285 observations missing.

var2 the auxiliary variable for the imputation model.

ExampleData10.csv *Simulated data for example 3*

Description

This is a simulated dataset as an example. The true value of theta is set to 1.0.

Usage

```
data1 <- read.csv("ExampleData10.csv", header=TRUE)
attach(data1)
```

Format

A data frame with 1000 observations on the following two variables.

`var1` the target variable for imputation with 285 observations missing.

`var2` the auxiliary variable for the imputation model.

<code>mranalyze2</code>	<i>R-function to analyze multiply-imputed datasets after multiple generalized ratio imputation</i>
-------------------------	--

Description

`mranalyze2` computes the mean, the standard error, the confidence intervals, the correlation coefficient, and the degrees of freedom after multiple generalized ratio imputation. It automatically combines M estimates.

Usage

```
mranalyze2(data, alpha = 0.05)
```

Arguments

<code>data</code>	Multiply-imputed data after <code>mrimpute2</code>
<code>alpha</code>	Significance level. Default is 0.05.

Value

<code>ybar</code>	Sample mean of the target variable
<code>se</code>	Standard error of the sample mean
<code>CI.LL</code>	Lower limit of the confidence interval
<code>CI.UL</code>	Upper limit of the confidence interval
<code>corr</code>	Correlation coefficient
<code>df</code>	Degrees of freedom

References

Takahashi, M. (2026) “Multiple generalized ratio imputation for missing data in official economic statistics: A flexible ratio estimator that automatically specifies the degree of heteroskedasticity.” Under review.

Takahashi, M. (2017a) “Multiple ratio imputation by the EMB algorithm: Theory and simulation.” *Journal of Modern Applied Statistical Methods*, 16 (1), 630-656.

Takahashi, M. (2017b) “JMASM44: Implementing multiple ratio imputation by the EMB Algorithm (R).” *Journal of Modern Applied Statistical Methods*, 16 (1), 657-673.

Example

```
outputimp <- read.csv(file.choose())
mranalyze2(outputimp)
mranalyze2(outputimp, alpha = 0.01)
```

mrdiag	<i>R-function to produce diagnostic plot for multiply-imputed datasets after multiple generalized ratio imputation</i>
--------	--

Description

mrdiag plots the density of observed data and multiply-imputed data.

Usage

```
mrdiag(data)
```

Arguments

data	Multiply-imputed data after mrimpute2
------	---------------------------------------

References

van Buuren, S. (2018) Flexible imputation.

Example

```
outputimp <- read.csv(file.choose())
mrdiag(outputimp)
```

mrimpute2	<i>R-function to compute multiple generalized ratio imputation</i>
-----------	--

Description

mrimpute2 multiply-imputes missing values based on multiple generalized ratio imputation.

Usage

```
mrimpute2(y, x, M = 5, iter = 5)
```

Arguments

<code>y</code>	Target variable for imputation
<code>x</code>	Auxiliary variable for the imputation model
<code>M</code>	Number of multiply-imputed datasets. Default is 5.
<code>iter</code>	Number of iterations to estimate the degree of heteroskedasticity. Default is 5.

References

Takahashi, M. (2026) “Multiple generalized ratio imputation for missing data in official economic statistics: A flexible ratio estimator that automatically specifies the degree of heteroskedasticity.” Under review.

Takahashi, M. (2017a) “Multiple ratio imputation by the EMB algorithm: Theory and simulation.” *Journal of Modern Applied Statistical Methods*, 16 (1), 630-656.

Takahashi, M. (2017b) “JMASM44: Implementing multiple ratio imputation by the EMB Algorithm (R).” *Journal of Modern Applied Statistical Methods*, 16 (1), 657-673.

Example

```
imp1 <- mrimpute2(var1, var2)
imp2 <- mrimpute2(var1, var2, M = 100)
imp3 <- mrimpute2(var1, var2, M = 100, iter = 10)
```

Value

A data frame with n observations on the following $M + 2$ variables.

V1 the target variable for imputation. The same as `var1`.

V2 the auxiliary variable for the imputation model. The same as `var2`.

V3 multiply-imputed data of `var1`, where $m = 1$.

V4 multiply-imputed data of `var1`, where $m = 2$.

⋮

VM+2 multiply-imputed data of `var1`, where $m = M$.

outputimp.csv	<i>Multiply-imputed data for example</i>
---------------	--

Description

This is an example of multiply-imputed dataset after multiple generalized ratio imputation by R function `mrimpute2`.

Usage

```
outputimp <- read.csv("outputimp.csv", header=TRUE)
attach(outputimp)
```

Format

A data frame with 1000 observations on the following six variables.

V1 the target variable for imputation. The same as `var1`.

V2 the auxiliary variable for the imputation model. The same as `var2`.

V3 multiply-imputed data of `var1`, where $m = 1$.

V4 multiply-imputed data of `var1`, where $m = 2$.

V5 multiply-imputed data of `var1`, where $m = 3$.

V6 multiply-imputed data of `var1`, where $m = 4$.

V7 multiply-imputed data of `var1`, where $m = 5$.

thetahat	<i>R-function to estimate the degree of heteroskedasticity</i>
----------	--

Description

`thetahat` estimates the degree of heteroskedasticity for multiple generalized ratio imputation.

Usage

```
thetahat(y, x, iter = 5, convplot = TRUE)
```

Arguments

<code>y</code>	Target variable for imputation
<code>x</code>	Auxiliary variable for the imputation model
<code>iter</code>	Number of iterations to estimate the degree of heteroskedasticity.

	Default is 5.
<code>convplot</code>	Convergence plot to check whether the estimated theta converges to a certain point. Default is TRUE.

Value

<code>theta.hat</code>	Estimated value of theta, which is the degree of heteroskedasticity.
------------------------	--

References

Takahashi, M. (2026) “Multiple generalized ratio imputation for missing data in official economic statistics: A flexible ratio estimator that automatically specifies the degree of heteroskedasticity.” Under review.

Example

```
t1 <- thetahat(var1, var2)
t1 <- thetahat(var1, var2, iter = 100)
t1 <- thetahat(var1, var2, iter = 100, convplot = FALSE)
```