

# Package ‘vsmi’

July 22, 2025

**Title** Variable Selection for Multiple Imputed Data

**Version** 0.1.0

**Description** Penalized weighted least-squares estimate for variable selection on correlated multiply imputed data and penalized estimating equations for generalized linear models with multiple imputation.

Reference:

Li, Y., Yang, H., Yu, H., Huang, H., Shen, Y\*. (2023) ``Penalized estimating equations for generalized linear models with multiple imputation", <doi:10.1214/22-AOAS1721>.

Li, Y., Yang, H., Yu, H., Huang, H., Shen, Y\*. (2023) ``Penalized weighted least-squares estimate for variable selection on correlated multiply imputed data", <doi:10.1093/jrsssc/qlad028>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.3.1

**Imports** MASS (>= 7.3-60), Matrix (>= 1.6-1.1), mice (>= 3.16.0), qif (>= 1.5)

**NeedsCompilation** no

**Author** Mingyue Zhang [aut],

Yang Li [aut],

Haoyu Yang [aut, cre]

**Maintainer** Haoyu Yang <haoyuyang@alu.ruc.edu.cn>

**Repository** CRAN

**Date/Publication** 2024-05-25 17:20:02 UTC

## Contents

generate_peek_missing_data . . . . .	2
generate_pwls_missing_data . . . . .	2
PEE . . . . .	3
PWLS . . . . .	4
vsmi . . . . .	5

<b>Index</b>	<b>7</b>
--------------	----------

---

```
generate_peek_missing_data
```

*Generate example data for PEE*

---

### Description

This is a function to generate example missing data for PEE

### Usage

```
generate_peek_missing_data(
  outcome = "binary",
  p = 20,
  n = 200,
  pt1 = 0.5,
  tbeta = c(3/4, (-3)/4, 3/4, (-3)/4, 3/4, (-3)/4, (-3)/4, 3/4),
  miss_sig = c(1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
)
```

### Arguments

outcome	The type of response variable Y, choose "binary" for binary response or "count" for poisson response,default "binary"
p	The dimension of the independent variable X,default 20.
n	The Number of rows of generated data,default 200.
pt1	Missing rate of independent variable X,default 0.5.
tbeta	True value of the coefficient,default c(3/4,(-3)/4,3/4,(-3)/4,3/4,(-3)/4,(-3)/4,3/4).
miss_sig	A 0-1 vector of length p, where 1 means that variable at the index is with missing,while 0 means that it without missing,default c(1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)

### Value

A Matrix,missing data with variables X in the first p columns and response Y at the last column.

---

```
generate_pwls_missing_data
```

*Generate example data for PWLS*

---

### Description

This is a function to generate example missing data for PWLS

**Usage**

```
generate_pwls_missing_data(
  p = 20,
  n = 200,
  pt1 = 0.5,
  pt2 = 0.5,
  tbeta = c(1, -1, 1, -1, 1, -1, -1, 1),
  miss_sig = c(0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0)
)
```

**Arguments**

p	The dimension of the independent variable X,default 20.
n	The Number of rows of generated data,default 200.
pt1	Missing rate of independent variable X,default 0.5.
pt2	Missing rate of response Y, default 0.5.
tbeta	True value of the coefficient,default c(1,-1,1,-1,1,-1,-1,1).
miss_sig	A 0-1 vector of length p, where 1 means that variable at the index is with missing,while 0 means that it without missing,default c(0,1,0,0,1,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0)

**Value**

A Matrix,missing data with variables X in the first p columns and response Y at the last column.

---

PEE	<i>Penalized estimating equations for generalized linear models with multiple imputation</i>
-----	--

---

**Description**

This is a function to impute missing data, estimate coefficients of generalized linear models and select variables for multiple imputed data sets, considering the correlation of multiple imputed observations.

**Usage**

```
PEE(
  misssdata,
  mice_time = 5,
  penalty,
  lamda.vec = seq(1, 4, length.out = 12),
  Gamma = c(0.5, 1, 1.5)
)
```

**Arguments**

misssdata	A Matrix,missing data with variables X in the first p columns and response Y at the last column.
mice_time	an integer, number of imputation.
penalty	The method for variable selection,choose from "lasso" or "alasso".
lamda.vec	Optimal tuning parameter for penalty,default seq(1,4,length.out=12).
Gamma	Parameter for adjustment of the Adaptive Weights vector in adaptive LASSO,default c(0.5,1,1.5).

**Value**

A Vsmi\_est object, contains estcoef and index\_sig , estcoef for estimate coefficients and index\_sig for selected variable index.

**Examples**

```
library(MASS)
library(mice)
library(qif)

data_with_missing <- generate_peek_missing_data(outcome="binary")
est.alasso <- PEE(data_with_missing,penalty="alasso")
est.lasso <- PEE(data_with_missing,penalty="lasso")

count_data_with_missing <- generate_peek_missing_data(outcome="count")
count_est.alasso <- PEE(data_with_missing,penalty="alasso")
count_est.lasso <- PEE(data_with_missing,penalty="lasso")
```

---

 PWLS

*Penalized weighted least-squares estimate for variable selection on correlated multiply imputed data*

---

**Description**

This is a functions to estimate coefficients of wighted least-squares model and select variables for multiple imputed data sets ,considering the correlation of multiple imputed observations.

**Usage**

```
PWLS(
  misssdata,
  mice_time = 5,
  penalty = "alasso",
  lamda.vec = seq(6, 24, length.out = 40),
  Gamma = c(0.5, 1, 2)
)
```

**Arguments**

misdata	A Matrix,missing data with variables X in the first p columns and response Y at the last column.
mice_time	An integer, number of imputation.
penalty	The method for variable selection,choose from "lasso" or "alasso".
lamda.vec	Optimal tuning parameter for penalty,default seq(1,4,length.out=12).
Gamma	Parameter for adjustment of the Adaptive Weights vector in adaptive LASSO,default c(0.5,1,1.5).

**Value**

A Vsmi\_est object, contains estcoef and index\_sig , estcoef for estimate coefficients and index\_sig for selected variable index.

**Examples**

```
library(MASS)
library(mice)
library(qif)
entire<-generate_pwls_missing_data()
est_lasso<-PWLS(entire,penalty="lasso")
est_alasso <- PWLS(entire,penalty = "alasso")
```

---

vsmi

*vsmi: Variable selection for multiple imputed data*


---

**Description**

This is a package to implementation penalized weighted least-squares estimate for variable selection on correlated multiply imputed data and penalized estimating equations for generalized linear models with multiple imputation.

**Functions**

[PEE](#):Penalized estimating equations for generalized linear models with multiple imputation

[PWLS](#) : Penalized weighted least-squares estimate for variable selection on correlated multiply imputed data

[generate\\_pwls\\_missing\\_data](#) : Generate example missing data for PWLS

[generate\\_pee\\_missing\\_data](#) : Generate example missing data for PEE

**Author(s)**

**Maintainer:** Haoyu Yang <haoyuyang@alu.ruc.edu.cn>

Authors:

- Mingyue Zhang
- Yang Li

# Index

`generate_peek_missing_data`, [2](#), [5](#)  
`generate_pwls_missing_data`, [2](#), [5](#)

PEE, [3](#), [5](#)  
PWLS, [4](#), [5](#)

`vsmi`, [5](#)  
`vsmi-package (vsmi)`, [5](#)