Package: AteMeVs (via r-universe)

August 30, 2024

Type Package

Title Average Treatment Effects with Measurement Error and Variable
Selection for Confounders

Version 0.1.0

Description A recent method proposed by Yi and Chen (2023)
<doi:10.1177/09622802221146308> is used to estimate the average
treatment effects using noisy data containing both measurement
error and spurious variables. The package 'AteMeVs' contains a
set of functions that provide a step-by-step estimation
procedure, including the correction of the measurement error
effects, variable selection for building the model used to
estimate the propensity scores, and estimation of the average
treatment effects. The functions contain multiple options for
users to implement, including different ways to correct for the
measurement error effects, distinct choices of penalty
functions to do variable selection, and various regression
models to characterize propensity scores.

License GPL-2

Encoding UTF-8

Imports MASS, ncvreg

NeedsCompilation yes

Author Li-Pang Chen [aut, cre], Grace Yi [aut]

Maintainer Li-Pang Chen <lchen723@nccu.edu.tw>

Date/Publication 2023-09-04 18:00:02 UTC

Repository https://lchen723.r-universe.dev

RemoteUrl https://github.com/cran/AteMeVs

RemoteRef HEAD

RemoteSha 7e7c6f533b02e5cdef11519a854fc1973724de4a

Contents

AteMeVs-package   *Estimation of average treatment effects with measurement error and variable selection for confounders*

Description

A recent method proposed by Yi and Chen (2023) <doi:10.1177/09622802221146308> is implemented to estimate the average treatment effects using noisy data containing both measurement error and spurious variables.

Details

The R package 'AteMeVs', which refers to estimation of the Average Treatment Effects with Measurement Error and Variable Selection for confounders, contains a set of functions that provide a step-by-step estimation procedure, including the correction of the measurement error effects, variable selection for building the model used to estimate the propensity scores, and estimation of the average treatment effects. The functions contain multiple options for users to implement, including different ways to correct for the measurement error effects, distinct choices of penalty functions to do variable selection, and various regression models to characterize propensity scores.

Author(s)

Chen, L.-P. and Yi, G. Y.

Maintainer: Li-Pang Chen <lchen723@nccu.edu.tw>

References

Yi, G. Y. and Chen, L.-P. (2023). Estimation of the average treatment effect with variable selection and measurement error simultaneously addressed for potential confounders. *Statistical Methods in Medical Research*, 32, 691-711.

DG *Generation of artificial data*

Description

This function is used to generate an artificial dataset, which contains potential outcomes, treatments, and error-prone confounders.

Usage

```
DG(X,Z,gamma_X,gamma_Z,Sigma_e,outcome="continuous")
```

Arguments

X	an $n \times p_x$ matrix of the error-prone confounders
Z	an $n \times p_z$ matrix of the precisely measured confounders
gamma_X	a p_x-dimensional vector of parameters corresponding to the error-prone confounders X
gamma_Z	a p_z-dimensional vector of parameters corresponding to the precisely measured confounders Z
Sigma_e	a $p_x \times p_x$ covariance matrix for the classical measurement error model
outcome	the indicator of the nature of the outcome variable; outcome="continuous" reflects normally distributed outcomes; outcome="binary" gives binary outcomes

Details

This function is used to generate artificial data, including potential outcomes, binary treatments, and error-prone and precisely measured confounders.

Value

data	an $n \times (2 + p_x + p_z)$ matrix of the artificial data. The first column is the potential outcome, and the second column is the binary treatment; column 3 to column (p_x+2) records error-prone confounders, and the remaining columns record precisely-measured confounders.

Author(s)

Chen, L.-P. and Yi, G. Y.

References

Yi, G. Y. and Chen, L.-P. (2023). Estimation of the average treatment effect with variable selection and measurement error simultaneously addressed for potential confounders. *Statistical Methods in Medical Research*, 32, 691-711.

Examples

```
library(MASS)
n = 800
p_x = 10       # dimension of parameters
p_z = 10
p = p_x + p_z
gamma_X = c(rep(1,2),rep(0,p_x-2))
gamma_Z = c(rep(1,2),rep(0,p_z-2))
gamma = c(gamma_X, gamma_Z)

mu_X = rep(0,p_x)
mu_Z = rep(0,p_z)

Sigma_X = diag(1,p_x,p_x)
Sigma_Z = diag(1,p_z,p_z)
Sigma_e = diag(0.2,p_x)
X = mvrnorm(n, mu_X, Sigma_X, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
Z = mvrnorm(n, mu_Z, Sigma_Z, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
data = DG(X,Z,gamma_X,gamma_Z,Sigma_e,outcome="continuous")
```

EST_ATE	*Estimation of the average treatment effect with the measurement error effects corrected and informative confounders accommodated*

Description

This function is used to estimate the average treatment effect by implementing the simulation and extrapolation (SIMEX) method with informative and error-eliminated confounders accommodated.

Usage

```
EST_ATE(data, PS="logistic", Psi=seq(0,1,length=10), K=200, gamma,p_x=p,
extrapolate="quadratic", Sigma_e, replicate = "FALSE",
RM = 0, bootstrap = 100)
```

Arguments

data
: an $n \times (p+2)$ matrix recording the data. The first column records the observed outcome, the second column displays the values for the binary treatment, and the remaining columns store the observed measurements for the confounders.

PS
: the specification of a link function in the treatment model. `logistic` refers to the logistic regression function, `probit` reflects the probit model, and `cloglog` gives the complementary log-log regression model.

Psi
: a user-specified sequence of non-negative values taken from an interval. The default is set as `Psi=seq(0,1,length=10)`.

p_x
: the dimension of the error-prone confounders

K
: a user-specified positive integer. The default is 200.

gamma	a vector of estimators for the treatment model, which is derived by using VSE_PS.
extrapolate	the extrapolation function in Step 3. `quadratic` reflects the quadratic polynomial function, `linear` gives the linear polynomial function, `RL` is the rational linear function, and `cubic` refers to the cubic polynomial function.
Sigma_e	the covariance matrix for the measurement error model
replicate	the indicator for the availability of repeated measurements in the confounders. `replicate = "FALSE"` refers no repeated measurements and `replicate = "TRUE"` indicates that repeated measurements exist in the dataset. The default is set as `replicate = "FALSE"`.
RM	a p_x-dimensional user-specified vector with entries being the number of repetitions for each confounder. For example, `RM = c(2,2,3)` indicates that three confounders in X have repeated measurements, where the first and second confounders have two repetitions and the third one has three repetitions. The default of RM is set as the p_x-dimensional zero vector, i.e., `RM = rep(0,p_x)`.
bootstrap	a user-specified positive integer representing the number of generated bootstrap samples to be applied with the estimation procedure

Details

This function is used to implement the simulation and extrapolation (SIMEX) method with informative confounders accommodated to to estimate the average treatment effect.

Value

estimate	a point estimate of the average treatment effect
variance	a variance estimate associated with the estimate of the average treatment effect
p-value	the resulting p-value of the average treatment effect

Author(s)

Chen, L.-P. and Yi, G. Y.

References

Yi, G. Y. and Chen, L.-P. (2023). Estimation of the average treatment effect with variable selection and measurement error simultaneously addressed for potential confounders. *Statistical Methods in Medical Research*, 32, 691-711.

See Also

SIMEX_EST, VSE_PS

Examples

```
library(MASS)
n = 800
p_x = 10      # dimension of parameters
p_z = 10
```

```
p = p_x + p_z
gamma_X = c(rep(1,2),rep(0,p_x-2))
gamma_Z = c(rep(1,2),rep(0,p_z-2))
gamma = c(gamma_X, gamma_Z)

mu_X = rep(0,p_x)
mu_Z = rep(0,p_z)

Sigma_X = diag(1,p_x,p_x)
Sigma_Z = diag(1,p_z,p_z)
Sigma_e = diag(0.2,p_x)
X = mvrnorm(n, mu_X, Sigma_X, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
Z = mvrnorm(n, mu_Z, Sigma_Z, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
data = DG(X,Z,gamma_X,gamma_Z,Sigma_e,outcome="continuous")


y = as.vector(SIMEX_EST(data,PS="logistic",Psi = seq(0,2,length=10),p_x=length(gamma_X),K=5,
    Sigma_e=diag(0.2,p_x)))
V = diag(1,length(y),length(y))

est_lasso_cv = VSE_PS(V,y,method="lasso",cv="TRUE")
EST_ATE(data, Psi = seq(0,2,length=10),p_x=length(gamma_X),K=5, gamma=est_lasso_cv,
        Sigma_e=diag(0.2,p_x),bootstrap = 10)

est_scad_cv = VSE_PS(V,y,method="scad",cv="TRUE")
EST_ATE(data, Psi = seq(0,2,length=10),p_x=length(gamma_X),K=5, gamma=est_scad_cv,
        Sigma_e=diag(0.2,p_x),bootstrap = 10)

est_mcp_cv = VSE_PS(V,y,method="mcp",cv="TRUE")
EST_ATE(data, Psi = seq(0,2,length=10),p_x=length(gamma_X),K=5, gamma=est_mcp_cv,
        Sigma_e=diag(0.2,p_x),bootstrap = 10)
```

SIMEX_EST   *Simulation and extrapolation (SIMEX) for the treatment model*

Description

This function employs the simulation and extrapolation (SIMEX) method to correct for the measurement error effects for confounders in the treatment model.

Usage

```
SIMEX_EST(data, PS="logistic", Psi=seq(0,1,length=10),p_x=p, K=200,
          extrapolate="quadratic", Sigma_e,
          replicate = "FALSE", RM = 0)
```

Arguments

data   an $n \times (p + 2)$ matrix of the data. The first column records the observed outcome, the second column displays the values for the binary treatment, and the remaining columns store the observed measurements for the confounders.

PS	the specification of a link function in the treatment model. `logistic` refers to the logistic regression function, `probit` reflects the probit model, and `cloglog` gives the complementary log-log regression model.
Psi	a user-specified sequence of non-negative values taken from an interval. The default is set as `Psi=seq(0,1,length=10)`.
p_x	the dimension of the error-prone confounders
K	a user-specified positive integer, with the default value set as 200
extrapolate	the extrapolation function in Step 3. `quadratic` reflects the quadratic polynomial function, `linear` gives the linear polynomial function, `RL` is the rational linear function, and `cubic` refers to the cubic polynomial function.
Sigma_e	the covariance matrix for the measurement error model
replicate	the indicator for the availability of repeated measurements in the confounders. `replicate = "FALSE"` refers no repeated measurements and `replicate = "TRUE"` indicates that repeated measurements exist in the dataset. The default is set as `replicate = "FALSE"`.
RM	a p_x-dimensional user-specified vector with entries being the number of repetitions for each confounder. For example, `RM = c(2,2,3)` indicates that three confounders in X have repeated measurements, where the first and second confounders have two repetitions and the third one has three repetitions. The default of RM is set as the p_x-dimensional zero vector, i.e., `RM = rep(0,p_x)`.

Details

This function is used to implement the simulation and extrapolation (SIMEX) method to estimate parameters in the treatment model.

Value

a vector of estimators in the treatment model

Author(s)

Chen, L.-P. and Yi, G. Y.

References

Yi, G. Y. and Chen, L.-P. (2023). Estimation of the average treatment effect with variable selection and measurement error simultaneously addressed for potential confounders. *Statistical Methods in Medical Research*, 32, 691-711.

Examples

```
library(MASS)
n = 800
p_x = 10      # dimension of parameters
p_z = 10
p = p_x + p_z
gamma_X = c(rep(1,2),rep(0,p_x-2))
```

```
gamma_Z = c(rep(1,2),rep(0,p_z-2))
gamma = c(gamma_X, gamma_Z)

mu_X = rep(0,p_x)
mu_Z = rep(0,p_z)

Sigma_X = diag(1,p_x,p_x)
Sigma_Z = diag(1,p_z,p_z)
Sigma_e = diag(0.2,p_x)
X = mvrnorm(n, mu_X, Sigma_X, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
Z = mvrnorm(n, mu_Z, Sigma_Z, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
data = DG(X,Z,gamma_X,gamma_Z,Sigma_e,outcome="continuous")


y = SIMEX_EST(data,PS="logistic",Psi = seq(0,2,length=10),p_x =length(gamma_X),
               K=5, Sigma_e=diag(0.2,p_x))
```

VSE_PS   *Variable selection for confounders*

Description

This function implements the penalized quadratic loss function to select the informative confounders.

Usage

```
VSE_PS(V,y,method="lasso",cv="TRUE",alpha=1)
```

Arguments

V	a user-specified matrix in the quadratic loss function
y	a vector determined by SIMEX_EST
method	it specifies a choice of the penalty function with options "lasso" (Tibshirani 1996), "scad" (Fan and Li 2001) and "mcp" (Zhang 2010). The default is set as method="lasso".
cv	the usage for choosing the tuning parameter. cv="TRUE" suggests the use of the cross-validation method, and cv="FALSE" allows the use of the BIC. The default is set as cv="TRUE".
alpha	the constant appearing in the Elastic Net penalty (Zou and Hastie 2005). The default value is 1.

Details

This function is used to do variable selection for informative confounders by various choices of penalty functions.

Value

a vector of estimators in the treatment model, where components with zero values represent confounders that are unimportant and need to excluded; components with nonzero values identify important confounders that enter the treatment model.

Author(s)

Chen, L.-P. and Yi, G. Y.

References

1. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360.

2. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267-288.

3. Yi, G. Y. and Chen, L.-P. (2023). Estimation of the average treatment effect with variable selection and measurement error simultaneously addressed for potential confounders. *Statistical Methods in Medical Research*, 32, 691-711.

4. Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38, 894-942.

5. Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67, 301-320.

See Also

SIMEX_EST

Examples

```
library(MASS)
n = 800
p_x = 10        # dimension of parameters
p_z = 10
p = p_x + p_z
gamma_X = c(rep(1,2),rep(0,p_x-2))
gamma_Z = c(rep(1,2),rep(0,p_z-2))
gamma = c(gamma_X, gamma_Z)

mu_X = rep(0,p_x)
mu_Z = rep(0,p_z)

Sigma_X = diag(1,p_x,p_x)
Sigma_Z = diag(1,p_z,p_z)
Sigma_e = diag(0.2,p_x)
X = mvrnorm(n, mu_X, Sigma_X, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
Z = mvrnorm(n, mu_Z, Sigma_Z, tol = 1e-6, empirical = FALSE, EISPACK = FALSE)
data = DG(X,Z,gamma_X,gamma_Z,Sigma_e,outcome="continuous")
```

```
y = as.vector(SIMEX_EST(data,PS="logistic",Psi = seq(0,2,length=10),p_x=length(gamma_X),
                K=5, Sigma_e=diag(0.2,p_x)))
V = diag(1,length(y),length(y))

VSE_PS(V,y,method="lasso",cv="TRUE")
VSE_PS(V,y,method="scad",cv="TRUE")
VSE_PS(V,y,method="mcp",cv="TRUE")
```

Index