Package: TVMR (via r-universe)

September 18, 2024

Type Package

Title Time-varying Mendelian randomization with time-continous modelling for the effect function

Version 0.1.0

Author Haodong Tian, Ashish Patel

Maintainer Haodong Tian <haodong.tian@mrc-bsu.cam.ac.uk>

Description Use functional principal components analysis (FPCA) within multivariable Mendelian randomization (MVMR) to estimate the time-varying effect function.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports stats, data.table, tidyverse, corrplot, parallel

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

RemoteUrl https://github.com/HDTian/TVMR

RemoteRef HEAD

RemoteSha 64706aec5a1c1494944a37888f737e2b8c2dbc5d

Contents

getX *simulate a data for the instrument and exposure data with time-varying variable information*

Description

simulate a data for the instrument and exposure data with time-varying variable information

Usage

```
getX(  N=10000,
       J=30,
       ZXmodel='A',
       MGX_used=NA,
       nSparse=10
       )
```

Arguments

N	a integer indicates the sample size
J	a integer indicates the number of genetic instrument
ZXmodel	a character indicates the form of the instrument-exposure model. Can be 'A', 'B', 'C', or 'D'
MGX_used	a matrix indicates the user-defined instrumental effect on the exposure. The default value is NA, corresponding to the case that the genetic effect is randomly valued.
nSparse	a integer indicates the number of the random measured timepoints for each individual.

Value

getX returns a list, which contains the information used for reproducing simulation, and the data matrix.

Examples

```
RES<-getX(J=30,ZXmodel='D')
```

`gmm_lm_onesample` *Multivariable GMM and Kleibergen's Lagrange multplier (LM) statistics with one-sample individual-level data*

Description

Multivariable GMM and Kleibergen's Lagrange multplier (LM) statistics with one-sample individual-level data

Usage

```
gmm_lm_onesample(X,
                 Y,
                 Z,
                 beta0=NA
                 )
```

Arguments

`Z`	n x J instrument matrix
`X`	n x K exposure matrix
`Y`	n x 1 outcome vector
`beta0`	the tested null of the causal parameter value (for LM test only)

Value

a result list, containing

`gmm_est`	K vector of causal effect estimates using GMM
`gmm_se`	K vector of standard errors corresponding to `gmm_est`
`variance_matrix`	K x K variance matrix corresponding to `gmm_est`
`gmm_pval`	K vector of p-values corresponding to gmm_est
`Q_stat`	overidentification test statistic
`Q_pval`	overidentification test p-value
`lm_stat`	the LM statistic value
`lm_pval`	p-value of the LM test of the null hypothesis H0: beta=beta0

Author(s)

Haodong Tian, Ashish Patel

IS *Instrumental strength with the conditional F statistics*

Description

Instrumental strength with the conditional F statistics

Usage

```
IS(J, K, timepoints, datafull)
```

Arguments

J	a integer indicates the number of instruments
K	a integer indicates the number of exposures
timepoints	a vector indicates the index of the exposures used for calculating the conditional F statistics
datafull	a data frame with the columns corresponding in order to the instruments, the exposures and the outcome

Value

It returns a table with the colunms of the coefficient of determination, F statistics, and the conditional F statistics for each exposures selected

MPCMR_GMM *Individual-level data time-varying MR*

Description

Multiple-principal-component Mendelian randomization fitting with GMM methods desgined for individual-level data. This function supports overlapping-data setting, where the overlapping data will first clumped to one-sample and continues the fitting afterward.

Usage

```
MPCMR_GMM( Gmatrix,
           res,
           Yvector,
           Gymatrix=NA,
           IDmatch=NA,
           nPC=NA,
           nL=NA,
           eigenfit=TRUE,
           polyfit=TRUE,
```

```
  LMCI=TRUE,
  LMCI2=TRUE,
  nLM=20,
  Parallel=TRUE,
  cores_used=NA,
  XYmodel=NA
)
```

Arguments

Gmatrix	a matrix indicates the instrument information
res	a list result of FPCA. res is usually derived by FPCA from the package fdapace. res must contains the following elements: res$cumFVE: a vector corresponds to the cumulative fraction-of-variance-explained (FVE). res$xiEst: a matrix contains the principal components values for each individual, where the rows correspond to the individuals and columns correspond to the peincipal compoemnts. res$workGrid: a vector indicates the working time Grid (such time points are mainly used for visulization purpose) res$phi: a matrix, where each column indicates the values of one eigenfunction over res$workGrid
Yvector	a vector indicates the individual point outcome
Gymatrix	a matrix indicates the instrument information for the outcome data. The default value is NA, which corresponds to the one-sample setting.
IDmatch	a vector indicates the overlapping data index. This is better generated by match(). The default value is NA, correponding to the one-sample setting. If you have an overlapping samples, and the ID vector for the exposure and outcome data are ID_X and ID_Y, you can get the IDmatch as myIDmatch <- match(ID_X, ID_Y).
nPC	a integer indicates the number of principal components used for MPCMR fitting. The default value is the number of principal compoments that just explain more than 95 percent variations.
nL	a integer indicates the degree of the polynomial (the number of polynomial basisfunctions)
eigenfit	logic. Whether to do MPCMR fitting with the eigenfunction as the basis function.
polyfit	logic. Whether to do MPCMR fitting with the polynomial set as the basis function.
LMCI	logic. Whether to calculate the CI with LM statistic where the basisfunction is the eigenfunction.
LMCI2	logic. Whether to calculate the CI with LM statistic where the basisfunction is the polynomial.
nLM	a integer indicates the number of increasing point for each dimension when calculating the CI with LM statistic.
Parallel	logic. Whether to use parallel computing. The default value is TRUE.
cores_used	an positive integer indicates how much cores will be used for the parallel computing. Only work when Parallel=TRUE. The default core number is the maximal cores number minus one.

`XYmodel` a character indicates the XY model. It should be only used for simulation purpose.

Details

Note that you should have individual-level data containing the genetic variants (i.e. genotype) information and the longitudinal information of the exposure of interest. Both information should be contained simultaneously for each individual.

The longitudinal information must contain both the exposure level and its corresponding measured time point (age). It allows for the exposure to be measured at different time points (ages) for every individual, and each individual can have a sparse measurement.

You will also have the outcome data, which can be summary-data or individual-data; one-sample or two-sample as the exposure data. Depending on the specific data setting, you can use different functions of TVMR. If you outcome data is individual-level data that is one-sample or overlapping sample with your exposure data, you just need to use `MPCMR_GMM`. If your outcome data is summary information or from two sample as the exposure data, you will need to use `MPCMR_GMM_twosample`.

Value

`MPCMR_GMM` retuns a list, consisting of various results, inclduing fitted parameters (and their standard errors), the weak IV assessment results, the IV validity assessment results, the fitted curve.

`nPC_used` how many principal components were used in MPCMR.

`L` the number of polynomial basis function used for fitting MPCMR.

`K` the number of eigenfunction basis function used for fitting MPCMR.

`ISres` the table results of the instrument strength. The columns are coefficient of determination, F value, conditional F, Q statistic value, degree-of-freedom of Q, and the p-value, respectively.

`scatterp` the MR scatter plot corresponding to the genetic association with the first and second principal compoments.

`one_sample_size` the sample size of the data finally used for MPCMR fitting.

`IV_validity_test` the IV validity test results, where the three values are the Q statistic, the degree-of-freedom and the p-value, respectively.

`MPCMRest` the fitted parameters for the eigenfunction basis set.

`MPCMRvar` the corresponding variance matrix of `MPCMRest`.

`p1` the fitted curve with eigenfunctin as the bsis function.

`ggdata1` the dataframe used for producing `p1` via ggplot.

`p2` the fitted curve with polynomial as the bsis function.

`ggdata2` the dataframe used for producing `p2` via ggplot.

`IV_validity_and_basisfunction_test` the IV validity test results considering the parameteric (polynomial) basis function, where the three values are the Q statistic, the degree-of-freedom and the p-value, respectively.

SE,MSE,Co,Coverage_rate,Co_LM,Coverage_rate_LM,sig_points,sig_points_LM gives some useful information when the true effect function is known (given by the argument XYmodel). They are used for simulation design.

The result name with the complementary symbol _p represents the results when the basis function are polynomial functions. For example, MPCMRest_p is the the fitted parameters for the polynmomial functions.

Author(s)

Haodong Tian

Examples

```
###see README.md file and TVMR/sim_real_illustration/MPCMR_illustration.R from GitHub
```

MPCMR_GMM_twosample *Two-sample time-varying MR fitting*

Description

Multiple-principal-component Mendelian randomization fitting with GMM methods. This function is desinged for the two-sample setting.

Usage

```
MPCMR_GMM_twosample(Gmatrix,
                    res,
                    by_used,
                    sy_used,
                    ny_used,
                    nPC=NA,
                    nL=NA,
                    eigenfit=TRUE,
                    polyfit=TRUE,
                    LMCI=TRUE,
                    LMCI2=TRUE,
                    nLM=20,
                    Parallel=TRUE,
                    cores_used=NA,
                    XYmodel=NA
                    )
```

Arguments

Gmatrix	a matrix indicates the instrument information

res	a list result of FPCA. res is usually derived by FPCA from the package fdapace. res must contains the following elements: res$cumFVE: a vector corresponds to the cumulative fraction-of-variance-explained (FVE). res$xiEst: a matrix contains the principal components values for each individual, where the rows correspond to the individuals and columns correspond to the peincipal compoemnts. res$workGrid: a vector indicates the working time Grid (such time points are mainly used for visulization purpose) res$phi: a matrix, where each column indicates the values of one eigenfunction over res$workGrid
by_used	a vector indicates the estimated genetic association with the outcome. The order of the genetic variants in by_used should be consistent with that order in Gmatrix
sy_used	a vector indicates the standard errors of the estimated genetic association with the outcome
ny_used	a integer indicates the smaple size of the outcome data
nPC	a integer indicates the number of principal components used for MPCMR fitting. The default value is the number of principal compoments that just explain more than 95 percent variations.
nL	a integer indicates the degree of the polynomial (the number of polynomial basisfunctions)
eigenfit	logic. Whether to do MPCMR fitting with the eigenfunction as the basis function.
polyfit	logic. Whether to do MPCMR fitting with the polynomial set as the basis function.
LMCI	logic. Whether to calculate the CI with LM statistic where the basisfunction is the eigenfunction.
LMCI2	logic. Whether to calculate the CI with LM statistic where the basisfunction is the polynomial.
nLM	a integer indicates the number of increasing point for each dimension when calculating the CI with LM statistic.
Parallel	logic. Whether to use parallel computing. The default value is TRUE and the cores used are the maximal cores munus one.
cores_used	an positive integer indicates how much cores will be used for the parallel computing. Only work when Parallel=TRUE. The default core number is the maximal cores number minus one.
XYmodel	a character indicates the XY model. It should be only used for simulation purpose.

Details

Note that you should have individual-level data containing the genetic variants (i.e. genotype) information and the longitudinal information of the exposure of interest. Both information should be contained simultaneously for each individual.

The longitudinal information must contain both the exposure level and its corresponding measured time point (age). It allows for the exposure to be measured at different time points (ages) for every individual, and each individual can have a sparse measurement.

If your individual outcome data is in two-sample with the exposure data or you just wish to treat your data as the two-sample case (e.g. your overlapping sample contains only a small fraction of identical individuals), then obtain the summary statistics from the individual outcome data, and then fit the MPCMR with summary outcome data.

Value

`MPCMR_GMM_twosample` retuns a list, consisting of various results, inclduing fitted parameters (and their standard errors), the weak IV assessment results, the IV validity assessment results, the fitted curve.

`nPC_used`	how many principal components were used in MPCMR.
`L`	the number of polynomial basis function used for fitting MPCMR.
`K`	the number of eigenfunction basis function used for fitting MPCMR.
`ISres`	the table results of the instrument strength. The columns are coefficient of determination, F value, conditional F, Q statistic value, degree-of-freedom of Q, and the p-value, respectively.
`scatterp`	the MR scatter plot corresponding to the genetic association with the first and second principal compoments.
`one_sample_size`	the sample size of the data finally used for MPCMR fitting.
`IV_validity_test`	the IV validity test results, where the three values are the Q statistic, the degree-of-freedom and the p-value, respectively.
`MPCMRest`	the fitted parameters for the eigenfunction basis set.
`MPCMRvar`	the corresponding variance matrix of `MPCMRest`.
`p1`	the fitted curve with eigenfunctin as the bsis function.
`ggdata1`	the dataframe used for producing `p1` via ggplot.
`p2`	the fitted curve with polynomial as the bsis function.
`ggdata2`	the dataframe used for producing `p2` via ggplot.
`IV_validity_and_basisfunction_test`	the IV validity test results considering the parameteric (polynomial) basis function, where the three values are the Q statistic, the degree-of-freedom and the p-value, respectively.

`SE`,`MSE`,`Co`,`Coverage_rate`,`Co_LM`,`Coverage_rate_LM`,`sig_points`,`sig_points_LM` gives some useful information when the true effect function is known (given by the argument `XYmodel`). They are used for simulation design.

The result name with the complementary symbol `_p` represents the results when the basis function are polynomial functions. For example, `MPCMRest_p` is the the fitted parameters for the polynmomial functions.

Author(s)

Haodong Tian

Examples

```
###see README.md file and TVMR/sim_real_illustration/MPCMR_illustration.R from GitHub
```

`plotEifun` *draw the eigenfunction plot based on a FPCA result*

Description

draw the eigenfunction plot based on a FPCA result

Usage

```
plotEifun(res)
```

Arguments

`res`	a list result of FPCA. `res` is usually derived by `FPCA` from the package fdapace. `res` must contains the following elements: `res$lambda`: the vector correpsonding the fraction of variance explained by each principal component (i.e. eigenvalues). `res$workGrid`: a vector indicates the working time Grid (such time points are mainly used for visulization purpose) `res$phi`: a matrix, where each column indicates the values of one eigenfunction over `res$workGrid`

Value

a gg-plot where each curve corresponds to one eigenfunction with the corresponding the eigenvalue (i.e. the fraction of variance explained by this eigenfunction).

`Qfunction` *Q function*

Description

obtain the Q statistic value and alos return the inference results based on Q statistic (e.g. weak-IV-robust estimation), based on an inputted effect parameter.

Usage

```
Qfunction(v,
          by,
          byse,
          B,
          BX,
          Sigma,
          Gam
          )
```

Arguments

v	a vector indicates the initial value of the effect parameter; length is L.
by	a vector indicates the genetic associations with the outcome; length is J.
byse	a vector indicates the standard error of by; length is J.
B	a transforming matrix; dimision is K x L. If using the full eigenfunction as the basisfunction, B is just the identity matrix I.
BX	a matrix corresponds to the genetic association with the (transformed) exposure; dimension is J x K.
Sigma	a array (K*K*J), each surface of which is the covariance matrix of the association estimators BX[j,].
Gam	a Gamma matrix, each row of which indicates the covariance of the estimated genetic association with the outcome and the estimated genetic assocation with the (transformed) exposure; dimension is J x K.

Value

return a list, containing

original_input_v	the original inputted value of the effect parameter.
inference_results	the Q-based (e.g. weak-IV-robust) estimation results, containing the parameter estimate and its standard errors.
Est	the estimated effext parameter.
Estvar	the variance matrix of Est
Qvalue	the Q satistic value

Index