Package 'gma'

September 19, 2017

Type Package

Title Granger Mediation Analysis

Version 1.0

Date 2018-08-23

Author Yi Zhao <zhaoyi1026@gmail.com>, Xi Luo <xi.rossi.luo@gmail.com>

Maintainer Yi Zhao <zhaoyi1026@gmail.com>

Description Performs Granger mediation analysis (GMA) for time series. This package includes a single level GMA model and a two-level GMA model, for time series with hierarchically nested structure. The single level GMA model for the time series of a single participant performs the causal mediation analysis which integrates the structural equation modeling and the Granger causality frameworks. A vector autoregressive model of order p is employed to account for the spatiotemporal dependencies in the data. Meanwhile, the model introduces the unmeasured confounding effect through a nonzero correlation parameter. Under the two-level model, by leveraging the variabilities across participants, the parameters are identifiable and consistently estimated based on a full conditional likelihood or a two-stage method. See Zhao, Y., & Luo, X. (2017), Granger Mediation Analysis of Multiple Time Series with an Application to fMRI, <arXiv:1709.05328> for details.

License GPL (>= 2)

Depends MASS, nlme, car

NeedsCompilation no

Repository CRAN

Date/Publication 2017-09-19 10:07:35 UTC

R topics documented:

---

gma-package   *Granger Mediation Analysis*

Description

gma performs Granger Mediation Analysis (GMA) of time series. This package includes a single level GMA model and a two-level GMA model, for time series with hierarchically nested structure. The single level GMA model for the time series of a single participant performs the causal mediation analysis which integrates the structural equation modeling and the Granger causality frameworks. A multivariate autoregressive model of order p is employed to account for the spatiotemporal dependencies in the data. Meanwhile, the model introduces the unmeasured confounding effect through a nonzero correlation parameter. Under the two-level model, by leveraging the variabilities across participants, the parameters are identifiable and consistently estimated using a two-stage method or a block coordinate descent method.

Details

Package:	gma
Type:	Package
Version:	1.0
Date:	2017-08-23
License:	GPL (>=2)

Author(s)

Yi Zhao <zhaoyi1026@gmail.com> and Xi Luo <xi.rossi.luo@gmail.com>

Maintainer: Yi Zhao <zhaoyi1026@gmail.com>

References

Zhao, Y., & Luo, X. (2017). *Granger Mediation Analysis of Multiple Time Series with an Application to fMRI.* arXiv preprint arXiv:1709.05328.

---

env.single   *Simulated single-level dataset*

Description

"env.single" is an R environment containing a data frame of data generated from 500 time points, and the true model parameters.

Usage

```
data("env.single")
```

Format

An R environment

data1 a data frame with Z the treatment assignment, M the mediator and R the interested outcome.

error1 a data frame with E1 and E2 the error time series of M and R, respectively.

theta a 3 by 1 vector, which is the coefficients (A,B,C) of the model.

Sigma a 2 by 2 matrix, which is the covariance matrix of two Gaussian white noise processes.

p the order of the vector autoregressive (VAR) model.

W a 2p by 2 matrix, which is the transition matrix of the VAR(p) model.

Delta a 2 by 2 matrix, which is the covariance matrix of the initial condition of the Gaussian white noise processes.

Details

The true parameters are set as follows. The number of time points is 500. The coefficients are set to be $A = 0.5$, $C = 0.5$ and $B = -1$. The variances of the model errors are $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ and the correlation is $\delta = 0.5$. For the VAR model, we consider the case $p = 1$, and the parameter settings satisfy the stationarity condition.

References

Zhao, Y., & Luo, X. (2017). *Granger Mediation Analysis of Multiple Time Series with an Application to fMRI.* arXiv preprint arXiv:1709.05328.

Examples

```
data(env.single)
dt<-get("data1",env.single)
```

---

env.two *Simulated two-level dataset*

Description

"env.two" is an R environment containing a data list generated from 50 subjects, and the parameter settings used to generate the data.

Usage

```
data("env.two")
```

Format

An R environment.

data2 a list of length 50, each contains a data frame with 3 variables.

error2 a list of length 50, each contains a data frame with 2 columns.

theta a 3 by 1 vector, which is the population level coefficients (A,B,C) of the model.

Sigma a 2 by 2 matrix, which is the covariance matrix of the two Gaussian white noise processes.

p the order of the vector autoregressive (VAR) model.

W a 2p by 2 matrix, which is the transition matrix of the VAR(p) model.

Delta a 2 by 2 matrix, which is the covariance matrix of the initial condition of the Gaussian white noise processes.

n a 50 by 1 matrix, is the number of time points for each subject.

Lambda the covariance matrix of the model errors in the coefficient regression model.

A a vector of length 50, is the A value in the single-level for each subject.

B a vector of length 50, is the B value in the single-level for each subject.

C a vector of length 50, is the C value in the single-level for each subject.

Details

The true parameters are set as follows. The number of subjects i $N = 50$. For each subject, the number of time points is a random draw from a Poisson distribution with mean 100. The population level coefficients are set to be $A = 0.5$, $C = 0.5$ and $B = -1$, and the variances of the Gaussian white noise process are assumed to be the same across participants with $\sigma_{1_i}^2 = 1$, $\sigma_{2_i}^2 = 4$ and the correlation is $\delta = 0.5$. For the VAR model, we consider the case $p = 1$, and the parameter settings satisfy the stationarity condition.

References

Zhao, Y., & Luo, X. (2017). *Granger Mediation Analysis of Multiple Time Series with an Application to fMRI.* arXiv preprint arXiv:1709.05328.

Examples

```
data(env.two)
dt<-get("data2",env.two)
```

gma	*Granger Mediation Analysis of Time Series*

Description

This function performs Granger Mediation Analysis (GMA) for time series data.

Usage

```
gma(dat, model.type = c("single", "twolevel"), method = c("HL", "TS", "HL-TS"),
  delta = NULL, p = 1, single.var.asmp = TRUE, sens.plot = FALSE,
  sens.delta = seq(-1, 1, by = 0.01), legend.pos = "topright",
  xlab = expression(delta), ylab = expression(hat(AB)), cex.lab = 1,
  cex.axis = 1, lgd.cex = 1, lgd.pt.cex = 1, plot.delta0 = TRUE,
  interval = c(-0.9, 0.9), tol = 1e-04, max.itr = 500, conf.level = 0.95,
  error.indep = TRUE, error.var.equal = FALSE,
  Sigma.update = TRUE, var.constraint = TRUE, ...)
```

Arguments

dat	a data frame or a list of data. When it is a data frame, it contains Z as the treatment/exposure assignment, M as the mediator and R as the interested outcome and model.type should be "single". Z, M and R are all in one column. When it is a list, the list length is the number of subjects. For a two-level dataset, each list contains one data frame with Z, M and R, and model.type should be "twolevel".
model.type	a character of model type, "single" for single level model, "twolevel" for two-level model.
method	a character of method that is used for the two-level GMA model. When delta is given, the method can be either "HL" (full likelihood) or "TS" (two-stage); when delta is not given, the method can be "HL", "TS" or "HL-TS". The "HL-TS" method estimates delta by the "HL" method first and uses the "TS" method to estimate the rest parameters.
delta	a number gives the correlation between the Gaussian white noise processes. Default value is NULL. When model.type = "single", the default will be 0. For two-level model, if delta = NULL, the value of delta will be estimated.
p	a number gives the order of the vector autoregressive model. Default value is 1.
single.var.asmp	
	a logic value indicates if in the single level model, the asymptotic variance will be used. Default value is TRUE.
sens.plot	a logic value. Default is FALSE. This is used only for single level model. When sens.plot = TRUE, the sensitivity analysis will be performed and plotted.
sens.delta	a sequence of delta values under which the sensitivity analysis is performed. Default is a sequence from -1 to 1 with increment 0.01. The elements with absolute value 1 will be excluded from the analysis.

legend.pos	a character indicates the location of the legend when `sens.plot = TRUE`. This is used for single level model.
xlab	a title for x axis in the sensitivity plot.
ylab	a title for y axis in the sensitivity plot.
cex.lab	the magnification to be used for x and y labels relative to the current setting of `cex`.
cex.axis	the magnification to be used for axis annotation relative to the current setting of `cex`.
lgd.cex	the magnification to be used for legend relative to the current setting of `cex`.
lgd.pt.cex	the magnification to be used for the points in legend relative to the current setting of `cex`.
plot.delta0	a logic value. Default is `TRUE`. When `plot.delta0 = TRUE`, the estimates when $\delta = 0$ is plotted.
interval	a vector of length two indicates the searching interval when estimating `delta`. Default is (-0.9,0.9).
tol	a number indicates the tolerance of convergence for the "HL" method. Default is `1e-4`.
max.itr	an integer indicates the maximum number of iteration for the "HL" method. Default is 500.
conf.level	a number indicates the significance level of the confidence intervals. Default is 0.95.
error.indep	a logic value. Default is `TRUE`. This is used for `model.type = "twolevel"`. When `error.indep = TRUE`, the error terms in the linear models for A, B and C are independent.
error.var.equal	
	a logic value. Default is `FALSE`. This is used for `model.type = "twolevel"`. When `error.var.equal = TRUE`, the variances of the error terms in the linear models for A, B and C are assumed to be identical.
Sigma.update	a logic value. Default is `TRUE`, and the estimated variances of the Gaussian white noise processes in the single level model will be updated in each iteration when running a two-level GMA model.
var.constraint	a logic value. Default is `TRUE`, and an interval constraint is added on the variance components in the higher level regression model of the two-level GMA model.
...	additional arguments to be passed.

Details

The single level GMA model is

$$M_t = Z_t A + E_{1t},$$

$$R_t = Z_t C + M_t B + E_{2t},$$

and for stochastic processes (E_{1t}, E_{2t}),

$$E_{1t} = \sum_{j=1}^{p} \omega_{11_j} E_{1,t-j} + \sum_{j=1}^{p} \omega_{21_j} E_{2,t-j} + \epsilon_{1t},$$

$$E_{2t} = \sum_{j=1}^{p} \omega_{12_j} E_{1,t-j} + \sum_{j=1}^{p} \omega_{22_j} E_{2,t-j} + \epsilon_{2t}.$$

A correlation between the Gaussian white noise $(\epsilon_{1t}, \epsilon_{2t})$ is assumed to be δ. The coefficients, as well as the transition matrix, are estimated by maximizing the conditional log-likelihood function. The confidence intervals of the coefficients are calculated based on the asymptotic joint distribution. The variance of AB estimator based on either the product method or the difference method is obtained from the Delta method. Under this single level model, δ is not identifiable. Sensitivity analysis for the indirect effect (AB) can be used to assess the deviation of the findings from assuming $\delta = 0$, when the independence assumption is violated.

The two-level GMA model is introduced to estimate δ from data without sensitivity analysis. It addresses the individual variation issue for datasets with hierarchically nested structure. For simplicity, we refer to the two levels of data by time series and subjects. Under the two-level GMA model, the data consists of N independent subjects and a time series of length n_i. The single level GMA model is first applied on the time series from a single subject. The coefficients then follow a linear model. Here we enforce the assumption that δ is a constant across subjects. The parameters are estimated through a full likelihood or a two-stage method.

Value

When model.type = "single",

Coefficients	point estimate of the coefficients, as well as the corresponding standard error and confidence interval. The indirect effect is estimated by both the produce (ABp) and the difference (ABd) methods.
D	point estimate of the causal coefficients in matrix form.
Sigma	estimate covariance matrix of the Gaussian white noise.
delta	the δ value used to estimate the rest parameters.
W	estimate of the transition matrix in the VAR(p) model.
LL	the conditional log-likelihood value.
time	the CPU time used, see system.time.

When model.type = "twolevel"

delta	the specified or estimated value of correlation parameter δ.
Coefficients	the estimated population level effect in the regression models.
Lambda	the estimated covariance matrix of the model errors in the higher level coefficient regression models.
Sigma	the estimated variances of ϵ_{1t} and ϵ_{2t} for each subject.
W	the estimated population level transition matrix.
HL	the value of full likelihood.
convergence	the logic value indicating if the method converges.
var.constraint	the interval constraints used for the variances in the higher level coefficient regression models.
time	the CPU time used, see system.time.

Author(s)

Yi Zhao, Brown University, <zhaoyi1026@gmail.com>; Xi Luo, Brown University, <xi.rossi.luo@gmail.com>.

References

Zhao, Y., & Luo, X. (2017). *Granger Mediation Analysis of Multiple Time Series with an Application to fMRI.* arXiv preprint arXiv:1709.05328.

Examples

```
# Example with simulated data

##############################################################################
# Single level GMA model
# Data was generated with 500 time points.
# The correlation between Gaussian white noise is 0.5.
data(env.single)
data.SL<-get("data1",env.single)

## Example 1: Given delta is 0.5.
gma(data.SL,model.type="single",delta=0.5)
# $Coefficients
#          Estimate          SE          LB          UB
# A      0.519090451 0.06048910  0.4005340  0.6376469
# C      0.487396067 0.12650909  0.2394428  0.7353493
# B     -0.951262962 0.07693595 -1.1020547 -0.8004713
# C2    -0.006395453 0.12125003 -0.2440411  0.2312502
# AB.p -0.493791520 0.07004222 -0.6310717 -0.3565113
# AB.d -0.493791520 0.17523161 -0.8372392 -0.1503439

## Example 2: Assume the white noise are independent.
gma(data.SL,model.type="single",delta=0)
# $Coefficients
#          Estimate          SE          LB          UB
# A      0.519090451 0.06048910  0.40053400 0.63764690
# C     -0.027668910 0.11136493 -0.24594015 0.19060234
# B      0.040982178 0.07693595 -0.10980952 0.19177387
# C2    -0.006395453 0.12125003 -0.24404115 0.23125024
# AB.p  0.021273457 0.04001358 -0.05715172 0.09969864
# AB.d  0.021273457 0.16463207 -0.30139946 0.34394638

## Example 3: Sensitivity analysis (given delta is 0.5)
# We comment out the example due to the computation time.
# gma(data.SL,model.type="single",delta=0.5,sens.plot=TRUE)
##############################################################################

##############################################################################
# Two-level GMA model
# Data was generated with 50 subjects.
# The correlation between white noise in the single level model is 0.5.
# The time series is generate from a VAR(1) model.
# We comment out our examples due to the computation time.
```

```
data(env.two)
data.TL<-get("data2",env.two)

## Example 1: Correlation is unknown and to be estimated.
# Assume errors in the coefficients model are independent.
# Add an interval constraint on the variance components.

# "HL" method
# gma(data.TL,model.type="twolevel",method="HL",p=1)
# $delta
# [1] 0.5176206
#
# $Coefficients
#            Estimate
# A         0.5587349
# C         0.7129338
# B        -1.0453097
# C2        0.1213349
# AB.prod -0.5840510
# AB.diff -0.5915989
#
# $time
#   user  system elapsed
# 12.285   0.381  12.684

# "TS" method
# gma(data.TL,model.type="twolevel",method="TS",p=1)
# $delta
# [1] 0.4993492
#
# $Coefficients
#            Estimate
# A         0.5569101
# C         0.6799228
# B        -0.9940383
# C2        0.1213349
# AB.prod -0.5535900
# AB.diff -0.5585879
#
# $time
#  user   system elapsed
# 7.745   0.175   7.934

## Example 2: Given the correlation is 0.5.
# Assume errors in the coefficients model are independent.
# Add an interval constraint on the variance components.

# "HL" method
# gma(data.TL,model.type="twolevel",method="HL",delta=0.5,p=1)
# $delta
# [1] 0.5
#
# $Coefficients
```

```
#          Estimate
# A        0.5586761
# C        0.6881703
# B       -0.9997898
# C2       0.1213349
# AB.prod -0.5585587
# AB.diff -0.5668355
#
# $time
#  user  system elapsed
# 0.889   0.023   0.913
###########################################################################
```

---

sim.data.ts.single   *Generate single-level simulation data*

Description

This function generates a single-level dataset with given parameters.

Usage

```
sim.data.ts.single(n, Z, A, B, C, Sigma, W, Delta = NULL, p = NULL, nburn = 100)
```

Arguments

n	an integer indicating the length of the time series.
Z	a vector of treatment/exposure assignment at each time point.
A	a numeric value of model coefficient.
B	a numeric value of model coefficient.
C	a numeric value of model coefficient.
Sigma	a 2 by 2 matrix, is the covariance matrix of the two Gaussian white noise processes.
W	a $2p$ by 2 matrix, is the transition matrix.
Delta	a 2 by 2 matrix, is the covariance matrix of the initial condition. Default is NULL, will be the same as Sigma.
p	a numeric value indicating the order of the vector autoregressive (VAR) model. Default is NULL, will be calculated based on W.
nburn	a integer indicating the number of burning sample. Default is 100.

Details

The single level GMA model is

$$M_t = Z_t A + E_{1t},$$

$$R_t = Z_t C + M_t B + E_{2t},$$

and for stochastic processes (E_{1t}, E_{2t}),

$$E_{1t} = \sum_{j=1}^{p} \omega_{11_j} E_{1,t-j} + \sum_{j=1}^{p} \omega_{21_j} E_{2,t-j} + \epsilon_{1t},$$

$$E_{2t} = \sum_{j=1}^{p} \omega_{12_j} E_{1,t-j} + \sum_{j=1}^{p} \omega_{22_j} E_{2,t-j} + \epsilon_{2t}.$$

`Sigma` is the covariance matrix of the Gaussian white noise $(\epsilon_{1t}, \epsilon_{2t})$, and `Delta` is the covariance matrix of $(\epsilon_{10}, \epsilon_{20})$. `W` is the transition matrix with element ω's.

Value

The function returns a list with two data frames. One is the data with variables `Z`, `M` and `R`; one is the data frame of (E_{1t}, E_{2t}).

Author(s)

Yi Zhao, Brown University, <zhaoyi1026@gmail.com>; Xi Luo, Brown University, <xi.rossi.luo@gmail.com>

References

Zhao, Y., & Luo, X. (2017). *Granger Mediation Analysis of Multiple Time Series with an Application to fMRI.* arXiv preprint arXiv:1709.05328.

Examples

```
###################################################
# Generate a single-level dataset

# covariance matrix of errors
delta<-0.5
Sigma<-matrix(c(1,2*delta,2*delta,4),2,2)

# model coefficients
A0<-0.5
B0<--1
C0<-0.5

# number of time points
n<-500

# generate a treatment assignment vector
set.seed(1000)
Z<-matrix(rbinom(n,size=1,prob=0.5),n,1)
```

```
# VAR(1) model
p<-1

# Delta and W matrices
Delta<-matrix(c(2,delta*sqrt(2*8),delta*sqrt(2*8),8),2,2)
W<-matrix(c(-0.809,0.154,-0.618,-0.5),2,2)

# number of burning samples
nburn<-1000

set.seed(1000)
data.single<-sim.data.ts.single(n,Z,A0,B0,C0,Sigma,W,Delta,p=p,nburn=nburn)
####################################################
```

---

sim.data.ts.two   *Generate two-level simulation data*

Description

This function generates a two-level dataset with given parameters.

Usage

```
sim.data.ts.two(Z.list, N, theta, Sigma, W, Delta = NULL, p = NULL,
  Lambda = diag(rep(1, 3)), nburn = 100)
```

Arguments

Z.list	a list of data. Each list is a vector containing the treatment/exposure assignment at each time point for the subject.
N	an integer, indicates the number of subjects.
theta	a vector of length 3, contains the population level causal effect coefficients.
Sigma	a 2 by 2 matrix, is the covariance matrix of the two Gaussian white noise processes in the single level model.
W	a $2p$ by 2 matrix, is the transition matrix in the single level model.
Delta	a 2 by 2 matrix, is the covariance matrix of the initial condition in the single level model. Default is NULL, will be the same as Sigma.
p	an integer, indicates the order of the vector autoregressive (VAR) model in the single level model. Default is NULL, will be calculated based on W.
Lambda	the covariance matrix of the model errors in the linear model of the model coefficients. Default is a 3 by 3 identity matrix.
nburn	a integer indicating the number of burning sample in the single level model. Default is 100.

Details

For the time series of length n_i of subject i, the single level GMA model is

$$M_{i_t} = Z_{i_t} A_i + E_{i_{1t}},$$

$$R_{i_t} = Z_{i_t} C_i + M_{i_t} B_i + E_{i_{2t}},$$

and

$$E_{i_{1t}} = \sum_{j=1}^{p} \omega_{i_{11_j}} E_{i_1, t-j} + \sum_{j=1}^{p} \omega_{i_{21_j}} E_{i_2, t-j} + \epsilon_{i_{1t}},$$

$$E_{i_{2t}} = \sum_{j=1}^{p} \omega_{i_{12_j}} E_{i_1, t-j} + \sum_{j=1}^{p} \omega_{i_{22_j}} E_{i_2, t-j} + \epsilon_{i_{2t}},$$

where `Sigma` is the covariance matrix of $(\epsilon_{i_{1t}}, \epsilon_{i_{2t}})$ (for simplicity, `Sigma` is the same across subjects). For coefficients A_i, B_i and C_i, we assume a multivariate regression model. The model errors are from a trivariate normal distribution with mean zero and covariance `Lambda`.

Value

data	a list of data. Each list is a data frame of (Z_t, M_t, R_t).
error	a list of data. Each list is a data frame of (E_{1t}, E_{2t}).
A	a vector of length N, the value of As.
B	a vector of length N, the value of Bs.
C	a vector of length N, the value of Cs.
type	a character indicates the type of the dataset.

Author(s)

Yi Zhao, Brown University, <zhaoyi1026@gmail.com>; Xi Luo, Brown University, <xi.rossi.luo@gmail.com>

References

Zhao, Y., & Luo, X. (2017). *Granger Mediation Analysis of Multiple Time Series with an Application to fMRI.* arXiv preprint arXiv:1709.05328.

Examples

```
####################################################
# Generate a two-level dataset

# covariance matrix of errors
delta<-0.5
Sigma<-matrix(c(1,2*delta,2*delta,4),2,2)

# model coefficients
A0<-0.5
B0<--1
C0<-0.5
```

```
theta<-c(A0,B0,C0)

# number of time points
N<-50
set.seed(2000)
n<-matrix(rpois(N,100),N,1)

# treatment assignment list
set.seed(1000)
Z.list<-list()
for(i in 1:N)
{
  Z.list[[i]]<-matrix(rbinom(n[i,1],size=1,prob=0.5),n[i,1],1)
}

# Lambda
Lambda<-diag(0.5,3)

# VAR(1) model
p<-1

# Delta and W matrices
Delta<-matrix(c(2,delta*sqrt(2*8),delta*sqrt(2*8),8),2,2)
W<-matrix(c(-0.809,0.154,-0.618,-0.5),2,2)

# number of burning samples
nburn<-1000

# set.seed(2000)
# data2<-sim.data.ts.two(Z.list,N,theta=theta,Sigma,W,Delta,p,Lambda,nburn)
####################################################
```

Index