Package 'ablasso'

May 3, 2024

Title Arellano-Bond LASSO Estimator for Dynamic Linear Panel Models

Version 1.0

Maintainer Junyu Chen <junyu.chen@outlook.de>

Description Implements the Arellano-Bond estimation method combined with LASSO for dynamic linear panel models. See Chernozhukov et al. (2024) ``Arellano-Bond LASSO Estimator for Dynamic Linear Panel Models''. arXiv preprint <doi:10.48550/arXiv.2402.00584>.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.1

Imports hdm, matrixStats, mvtnorm, stats

Depends R (>= 2.10)

LazyData true

NeedsCompilation no

Author Victor Chernozhukov [aut],
Ivan Fernandez-Val [aut],
Chen Huang [aut],
Weining Wang [aut],
Junyu Chen [cre]

Repository CRAN

Date/Publication 2024-05-03 13:00:13 UTC

R topics documented:

ablasso_mv_ss *AB-LASSO Estimator with Random Sample Splitting for Multivariate Models*

Description

Implements the AB-LASSO estimation method for the multivariate model $Y_{it} = \alpha_i + \gamma_t + \sum_{j=1}^{L} \beta_j Y_{i,t-j} + \theta_0 D_{it} + \theta_1 C_{i,t-1} + \varepsilon_{it}$, with random sample splitting. Note that D_{it} and C_{it} are predetermined with respect to ε_{it}.

Usage

```
ablasso_mv_ss(Y, D, C, lag = 1, Kf = 2, nboot = 100, seed = 202302)
```

Arguments

Y	A P x N (number of time periods x number of individuals) matrix containing the outcome/response variable Y.
D	A P x N (number of time periods x number of individuals) matrix containing the policy variable/treatment D.
C	A list of P x N matrices containing other treatments and control variables.
lag	The lag order of Y_{it} included in the covariates, default is 1.
Kf	The number of folds for K-fold cross-validation, with options being 2 or 5, default is 2.
nboot	The number of random sample splits, default is 100.
seed	Seed for random number generation, default 202302.

Value

A dataframe that includes the estimated coefficients ($\beta_j, \theta_0, \theta_1$), their standard errors, and T-statistics.

Examples

```
# Use the Covid data
N = length(unique(covid_data$fips))
P = length(unique(covid_data$week))
Y = matrix(covid_data$logdc, nrow = P, ncol = N)
D = matrix(covid_data$dlogtests, nrow = P, ncol = N)
C = list()
C[[1]] = matrix(covid_data$school, nrow = P, ncol = N)
C[[2]] = matrix(covid_data$college, nrow = P, ncol = N)
C[[3]] = matrix(covid_data$pmask, nrow = P, ncol = N)
C[[4]] = matrix(covid_data$pshelter, nrow = P, ncol = N)
C[[5]] = matrix(covid_data$pgather50, nrow = P, ncol = N)

results.kf2 <- ablasso_mv_ss(Y = Y, D = D, C = C, lag = 4, nboot = 2)
```

```
print(results.kf2)
results.kf5 <- ablasso_mv_ss(Y = Y, D = D, C = C, lag = 4, Kf = 5, nboot = 2)
print(results.kf5)
```

ablasso_uv *AB-LASSO Estimator Without Sample Splitting*

Description

Implements the AB-LASSO estimation method for the univariate model $Y_{it} = \alpha_i + \gamma_t + \theta_1 Y_{i,t-1} + \theta_2 D_{it} + \varepsilon_{it}$, without sample splitting. Note that D_{it} is predetermined with respect to ε_{it}.

Usage

```
ablasso_uv(Y, D)
```

Arguments

Y	A P x N (number of time periods x number of individuals) matrix containing the outcome/response variable Y.
D	A P x N (number of time periods x number of individuals) matrix containing the policy variable/treatment D.

Value

A list with three elements:

- theta.hat: Estimated coefficients.
- std.hat: Estimated Standard errors.
- stat: T-Statistics.

Examples

```
# Generate data
data1 <- generate_data(N = 300, P = 40)

# You can use your own data by providing matrices `Y` and `D`
results <- ablasso_uv(Y = data1$Y, D = data1$D)
print(results)
```

ablasso_uv_ss *AB-LASSO Estimator with Random Sample Splitting*

Description

Implements the AB-LASSO estimation method for the univariate model $Y_{it} = \alpha_i + \gamma_t + \theta_1 Y_{i,t-1} + \theta_2 D_{it} + \varepsilon_{it}$, incorporating random sample splitting. Note that D_{it} is predetermined with respect to ε_{it}.

Usage

```
ablasso_uv_ss(Y, D, nboot = 100, Kf = 2, seed = 202304)
```

Arguments

`Y` A `P` x `N` (number of time periods x number of individuals) matrix containing the outcome/response variable variable `Y`.

`D` A `P` x `N` (number of time periods x number of individuals) matrix containing the policy variable/treatment `D`.

`nboot` The number of random sample splits, default is `100`.

`Kf` The number of folds for K-fold cross-validation, with options being `2` or `5`, default is `2`.

`seed` Seed for random number generation, default `202304`.

Value

A list with three elements:

- theta.hat: Estimated coefficients.
- std.hat: Estimated Standard errors.
- stat: T-Statistics.

Examples

```
# Generate data
data1 <- generate_data(N = 300, P = 40)

# You can use your own data by providing matrices `Y` and `D`
results.ss <- ablasso_uv_ss(Y = data1$Y, D = data1$D, nboot = 2)
print(results.ss)

results.ss2 <- ablasso_uv_ss(Y = data1$Y, D = data1$D, nboot = 2, Kf = 5)
print(results.ss2)
```

covid_data *COVID-19 Spread and School Policy Effects Data*

Description

A balanced panel data set analyzing the impact of K-12 school openings and other policy measures on the spread of COVID-19 across U.S. counties. The data spans 32 weeks from April 1st to December 2nd, 2020, and covers 2510 counties.

Usage

```
covid_data
```

Format

A data frame with 80320 (2510 counties times 32 weeks) rows and 9 columns. Each column represents a variable:

fips County FIPS

week Week

school A measure of visits to K-12 schools from SafeGraph foot traffic data

logdc Logarithm of the number of reported COVID-19 cases

pmask Policy indicators on mask mandates

pgather50 Policy indicators on ban on gatherings of more than 50 persons

college Measure of visits to colleges

pshelter Policy indicators on stay-at-home orders

dlogtests A measure of the weekly growth rate in the number of tests

Source

Data initially provided by Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf on the GitHub repository `https://github.com/ubcecon/covid-schools`. Counties with missing values are dropped to obtain a balanced panel dataset.

Examples

```
data(covid_data) # Access the dataset
```

generate_data *Generate a Dataset for Simulations*

Description

Generates data according to the following process: $Y_{it} = \alpha_i + \gamma_t + \theta_1 Y_{i,t-1} + \theta_2 D_{it} + \varepsilon_{it}$ and $D_{it} = \rho D_{i,t-1} + v_{i,t}$. Note that D_{it} is predetermined with respect to ε_{it}.

Usage

```
generate_data(
  N,
  P,
  sigma_alpha = 1,
  sigma_gamma = 1,
  sigma_eps.d = 1,
  sigma_eps.y = 1,
  cov_eps = 0.5,
  rho = 0.5,
  theta = c(0.8, 1),
  seed = 202304
)
```

Arguments

N	An integer specifying the number of individuals.
P	An integer specifying the number of time periods.
sigma_alpha	Standard deviation for the normal distribution from which the individual effect alpha is drawn; default is 1.
sigma_gamma	Standard deviation for the normal distribution from which the time effect gamma is drawn; default is 1.
sigma_eps.d	Standard deviation for the error term associated with the policy variable/treatment (D); default is 1.
sigma_eps.y	Standard deviation for the error term associated with the outcome/response variable (Y); default is 1.
cov_eps	Covariance between error terms of Y and D, default 0.5.
rho	Autocorrelation coefficient for D across time, default 0.5.
theta	Regression Coefficients for univariate AR(1) dynamic panal, default c(0.8, 1).
seed	Seed for random number generation, default 202304.

Value

A list of two P x N matrices named Y (outcome/response variable) and D (policy variable/treatment).

Examples

```
# Generate data using default parameters
data1 <- generate_data(N = 300, P = 40)
str(data1)

data2 <- generate_data(N = 500, P = 20)
str(data2)
```

Index