Package 'ClusterVAR'

July 21, 2025

Type Package

Title Fitting Latent Class Vector-Autoregressive (VAR) Models

Version 0.0.8

Maintainer Anja Ernst <a.f.ernst@rug.nl>

Description Estimates latent class vector-autoregressive models via EM algorithm on time-series data for model-based clustering and classification. Includes model selection criteria for selecting the number of lags and clusters.

BugReports https://github.com/anieBee/ClusterVAR/issues

Encoding UTF-8

LazyData true

License GPL-2

Imports stats, utils, graphics, grDevices, fastDummies, MASS, mvtnorm, scales, foreach, parallel, doParallel, parabar, iterators

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Anja Ernst [aut, cre],
Jonas Haslbeck [aut]

Repository CRAN

Date/Publication 2024-12-09 13:40:25 UTC

Contents

1

---

coef.ClusterVAR *Show coefficients of given Model*

Description

Extracts the coefficients of a given model from the output object of the LCVAR function.

Usage

```
## S3 method for class 'ClusterVAR'
coef(object, Model,...)
```

Arguments

object	An output object of the LCVAR function.
Model	An integer vector specifying the model for which coefficients should be shown. For example, Model = c(1,1,2) returns the coefficients of a model with three clusters (i.e., latent classes), in which the first cluster has one lag, the second cluster has 1 lag, and the third cluster has 2 lags. The vector length indicates the number of clusters in the model of interest.
...	Pass additional arguments.

Value

Lags	An integer vector specifying for which model the coefficients are shown. For example, Lags = c(1,1,2) indicates a model with three clusters, in which the first cluster has one lag, the second cluster has 1 lag, and the third cluster has 2 lags.
Classification	The crisp classification for each individual into a cluster. Crisp classifications are made based on an individual's modal cluster membership probabilities.
VAR_coefficients	
	The cluster-wise vector-autoregressive coefficients for each cluster.
Exogenous_coefficients	
	The cluster-wise exogenous coefficients for each cluster. The first column in each array indicates the (conditional) within-person mean in that cluster. If exogenous variable(s) were specified, the other columns indicate the influences of the exogenous variable(s) in that cluster.
Sigma	The cluster-wise innovation covariance matrix for each cluster.
Proportions	The mixing proportions for each cluster. These can be considered as the proportion of individuals belonging to the respective cluster.
PredictableTimepoints	
	The total number of time-points in this dataset that could be predicted because the previous time-point(s) were observed. In case of unequal lags across different clusters, the number of time-points for each person are weighted by their posterior cluster-membership probability. See also numberPredictableObservations().
Converged	A logical value that indicates whether this model converged.

Author(s)

Anja Ernst & Jonas Haslbeck

Examples

```
LCVAR_outExample <- LCVAR(Data = ExampleData,
                          yVars = 1:4, ID = 5, Beep = 6,
                          xContinuous = 7, xFactor = 8,
                          Clusters = 2, Lags = 1:2, smallestClN = 3,
                          Cores = 2, RndSeed = 3, Rand = 2,
                          it = 25)

coef(LCVAR_outExample, Model = c(1, 1))
coef(LCVAR_outExample, Model = c(2, 2))
```

LCVAR   *Fitting Latent Class VAR Models*

Description

Function to fit a Latent Class VAR model with a given number of latent classes.

Usage

```
LCVAR(Data, yVars, Beep, Day = NULL, ID,
            xContinuous = NULL, xFactor = NULL,
            Clusters, Lags, Center = FALSE,
            smallestClN = 3, Cores = 2,
            RndSeed = NULL, Rand = 50, Rational = TRUE,
            Initialization = NULL, SigmaIncrease = 10,
            it = 50, Conv = 1e-05, pbar = TRUE, verbose = TRUE,
            Covariates = "equal-within-clusters", ...)
```

Arguments

Data	The data provided in a data.frame.
yVars	An integer vector specifying the position of the column(s) in dataframe `Data` that contain the endogenous variables (= the VAR time series).
Beep	An integer specifying the position of the column in dataframe `Data` that contains the time-point.
Day	Optional argument. An integer specifying the position of the column in dataframe `Data` that contains the variable that indicates the day of measurement. If `Day` is supplied here, measurements on the previous day are not used to predict measurements on the current day. Instead, the first `Lags` observations within each day are excluded from the calculation of VAR coefficients.

ID	An integer specifying the position of the column in dataframe `Data` that contains the ID variable for every participant.
xContinuous	Optional argument. An integer vector specifying the position of the column(s) in dataframe Data that contain the continuous exogenous variable(s), if present. Exogenous variables are also known as covariates or as moderators for the within-person mean.
xFactor	Optional argument. An integer vector specifying the position of the column(s) in dataframe Data that contain the categorical exogenous variable(s), if present. Exogenous variables are also known as covariates or as moderators for the within-person mean.
Clusters	An integer or integer vector specifying the numbers of latent classes (i.e., clusters) for which LCVAR models are to be calculated.
Lags	An integer or integer vector specifying the number of VAR(p) lags to consider. Needs to be a sequence of subsequent integers. The maximum number supported is `Lags = 3`.
Center	Logical, indicating whether the data (i.e., the endogenous variables) should be centered per person before calculations. If `Center = TRUE`, the differences in within-person means are removed from the data and the clustering is based only on similarity in VAR coefficients and (if exogenous variable(s) are specified) similarities in infleunces of exogenous variable(s). If `Center = FALSE`, the clustering is also based on similarities in within-person means. Defaults to `Center = FALSE`.
smallestClN	An integer specifying the lowest number of individuals allowed in a cluster. When during estimation the crisp cluster membership of a cluster indicates less than `smallestClN` individuals, the covariance matrix and the posterior probabilities of cluster membership are reset. Defaults to `smallestClN = 3`.
Cores	A positive integer specifying the number of cores used to parallelize the computations. Specifying a high number of available cores can speed up computation. Defaults to `Cores = 2`.
RndSeed	Optional argument. An integer specifying the value supplied to `set.seed()`, which guarantees reproducible results. If not specified, no seed is set.
Rand	The number of pseudo-random EM-starts used in fitting each possible model. For pseudo-random starts K individuals are randomly selected as cluster centres. Then individuals are partitioned into the cluster to which their individual VAR and individual covariate coefficients are closest. High numbers (e.g., 50 and above) ensure that a global optimum will be found, but will take longer to compute. Defaults to `Rand = 50`.
Rational	Logical, indicating whether a rational EM-start should be used in addition to the other EM-starts. Defaults to `Rational = TRUE`. Rational starts are based on the k-means partitioning of individuals' ideographic VAR and ideographic covariate coefficients.
Initialization	Optional argument. An integer specifying the position of a column in dataframe `Data` that contains a guess at participants' cluster membership for a fixed number of clusters, if available. This initialization will be used as an additional EM-start.

SigmaIncrease	A numerical value specifying the value by which every element of Sigma will be increased when posterior probabilities of cluster memberships are reset. Defaults to `SigmaIncrease = 10`.
it	An integer specifying the maximum number of EM-iterations allowed for every EM-start. After completing `it` EM-iterations, an EM-start is forced to terminate. High numbers (e.g., 100 and above) ensure convergence, but will take longer to compute. Defaults to `it = 50`.
Conv	A numerical value specifying the convergence criterion of the log likelihood to determine convergence of an EM-start. For details see Ernst et al. (2020) Inter-individual differences in multivariate time series: Latent class vector-autoregressive modelling. Defaults to `Conv = 1e-05`.
pbar	If `pbar = TRUE`, a progress bar is shown. Defaults to `pbar = TRUE`.
verbose	If `verbose = FALSE`, output messages are limited. Additionally, the `pbar` argument is overridden, so the progress bar is not printed. Defaults to `verbose = TRUE`.
Covariates	Constraints on the parameters of the exogenous variable(s). So far only `Covariates = "equal-within-clusters"` can be specified.
...	Additional arguments passed to the function.

Details

This function estimates the latent class vector-autoregressive model to obtain latent classes (i.e., clusters) of individuals who are similar in VAR coefficients and (if specified) in within-person means and infleunces of exogenous variable(s).

$$y_{i,t} = w_{i,t} + \mu_k + \beta_k x_{i,t}$$

$$w_{i,t} = \left(\sum_{a=1}^{p} \Phi_{k,a} w_{i,t-a} \right) + u_{i,t} \qquad u_{i,t} \sim N(0, \Sigma_k)$$

Here μ_k represents an m x 1 vector that contains the cluster-wise conditional within-person mean for each y-variable in cluster k. β_k represents an m x q matrix that expresses the cluster-wise moderating influence of q exogenous variables ($x_{i,t}$) on the within-person means in cluster k. $\Phi_{k,a}$ represents an m×m matrix containing the cluster-wise VAR coefficients at lag a for cluster k. See the references below for details.

Value

An object of class 'ClusterVAR' providing several LCVAR models. The details of the output components are as follows:

Call	A list of arguments from the original function call.
All_Models	All LCVAR models across all number of clusters, lag combinations, and number of EM-starts. `All_Models[[a]][[b]][[c]]` contains all information for the LCVAR model for the ath number of clusters that was specified in `Clusters`, for the bth combination of lag orders, based on the combination of lags that was

specified in Lags, on the cth EM-start. To find the ideal model across all of them
use summary(), to view the coefficients of a given model, use coef().

Runtime   The runtime the function took to complete.

Author(s)

Anja Ernst

References

Ernst, A. F., Albers, C. J., Jeronimus, B. F., & Timmerman, M. E. (2020). Inter-individual differences in multivariate time-series: Latent class vector-autoregressive modeling. European Journal of Psychological Assessment, 36(3), 482–491. doi:10.1027/10155759/a000578

Examples

```
head(SyntheticData)
LCVAR_outExample1 <- LCVAR(Data = SyntheticData,
                           yVars = 1:4, ID = 5, Beep = 9, Day = 10,
                           xContinuous = 7, xFactor = 8,
                           Clusters = 1:2, Lags = 1,
                           Center = TRUE,
                           Cores = 2, # Adapt to local machine
                           RndSeed = 123, Rand = 1, it = 25)
summary(LCVAR_outExample1)
summary(object = LCVAR_outExample1, show = "GNL", Number_of_Lags = 1)
coef(LCVAR_outExample1, Model = c(1, 1))


head(ExampleData)
LCVAR_outExample2 <- LCVAR(Data = ExampleData,
                           yVars = 1:4, ID = 5, Beep = 6,
                           xContinuous = 7, xFactor = 8,
                           Clusters = 1:2, Lags = 1:2,
                           Center = FALSE,
                           Cores = 2, RndSeed = 123,
                           Rand = 1,
                           it = 25, Conv = 1e-05)
summary(LCVAR_outExample2)
summary(object = LCVAR_outExample2, show = "GNL", Number_of_Lags = 1)
summary(object = LCVAR_outExample2, show = "GNC", Number_of_Clusters = 2)
coef(LCVAR_outExample2, Model = c(1, 1))
plot(LCVAR_outExample2, show = "specific", Model = c(1, 1))


LCVAR_outExample3 <- LCVAR(Data = ExampleData,
                           yVars = c("Item1", "Item2","Item3", "Item4"),
                           ID = "Person", Beep = "Timepoint",
                           xContinuous = "ContiniousVariable",
                           xFactor = "CategoricalVariable",
```

```
                               Clusters = 1:2, Lags = 1:2,
                               Center = FALSE,
                               Cores = 2, RndSeed = 123,
                               Rand = 1,
                               it = 25, Conv = 1e-05)
     plot(LCVAR_outExample3, show = "GNL", Number_of_Lags = 1)
     plot(LCVAR_outExample3, show = "GNC", Number_of_Clusters = 2)
```

numberPredictableObservations

Determine the number of observations that can be predicted

Description

numberPredictableObservations is a function to determine the number of observations in a given dataset that can be predicted based on the availability of previous observations, considering a specified time-lag.

Usage

```
numberPredictableObservations(Data, yVars, Beep, Day = NULL, ID,
                              xContinuous = NULL, xFactor = NULL, Lags, ...)
```

Arguments

Data	The data provided in a data.frame.
yVars	An integer vector specifying the position of the column(s) in dataframe Data that contain the endogenous variables (= the VAR time series).
Beep	An integer specifying the position of the column in dataframe Data that contains the time-point.
Day	Optional. An integer specifying the position of the column in dataframe Data that contains the variable that indicates the day of measurement. If Day is supplied here, measurements on the previous day are not used to predict measurements on the current day. Instead, the first Lags observations within each day are excluded from the calculation of VAR coefficients.
ID	An integer specifying the position of the column in dataframe Data that contains the ID variable for every participant.
xContinuous	Optional argument. An integer vector specifying the position of the column(s) in dataframe Data that contain the continuous exogenous variable(s), if present. Exogenous variables are also known as covariates or as moderators for the within-person mean.

xFactor	Optional argument. An integer vector specifying the position of the column(s) in dataframe Data that contain the categorical exogenous variable(s), if present. Exogenous variables are also known as covariates or as moderators for the within-person mean.
Lags	An integer or integer vector specifying the number of VAR(p) lags to consider. Needs to be a sequence of subsequent integers. The maximum number supported is Lags = 3.
...	Additional arguments passed to the function.

Details

This function determines the number of observations in a given dataset that can be predicted based on previous observations. For instance, in a lag-1 model, if an observation is missing, the observation at the next time-point cannot be predicted. Similarly, in a lag-2 model, if an observation is missing, the observations at the next two time-points cannot be predicted. The output gives the number of predictable observations for each of the endogenous variables that was specified under yVars. The number of predictable observations is the same for all endogenous variables.

Value

Predictable observations per subject

> The number of predictable observations for each endogenous variable per subject, considering a specified time-lag.

Total predictable observations

> The total number of predictable observations summed over all subjects in the dataset for each endogenous variable, considering a specified time-lag.

Author(s)

Anja Ernst & Jonas Haslbeck

Examples

```
head(SyntheticData)

Obs <- numberPredictableObservations(Data = SyntheticData, yVars = 1:4,
                                     Beep = 9, Day = 10,  ID = 5, Lags = 1:3)

Obs

Obs$`Predictable observations per subject`$`1 Lag`
```

---

plot.ClusterVAR   *Visualizing Model fit and information criteria*

Description

Creates a variety of plots summarizing fitted LCVAR models.

Usage

```
## S3 method for class 'ClusterVAR'
plot(x,
    show,
    Number_of_Clusters = NULL,
    Number_of_Lags = NULL,
    Model = NULL,
    mar_heat = c(2.5,2.5,2,1),
    ...)
```

Arguments

x	An output object of the `ClusterVAR` function
show	Indicate summaries to plot. `show = "GNC"` compares models with different lags for a given number of clusters, specified with `Number_of_Clusters`. `show = "GNL"` compares models with different number of clusters, given the a number of lags specified with `Number_of_Lags`.
	Alternatively, the VAR matrices of a specific model can be visualized. `show = specific` visualizes the VAR matrices across clusters for a given model which is specified with the argument `Model`. `show = specificDiff` displays the pairwise differences between clusters for a specific model, also specified with the argument `Model`. `show = specific` and `show = specificDiff` are currently only implemented for models with 1 lag.
Number_of_Clusters	An integer. Specify the fixed number of clusters when using `show = "GNC"`. Defaults to `Number_of_Clusters = NULL`
Number_of_Lags	An integer. Specify the fixed number of lags when using `show = "GNL"`. Defaults to `Number_of_Lags = ` the lowest number of lags specified in object.
Model	An integer vector. Specify when using `show = "specific"` or `show = specificDef`. Indicates the model for which coefficients should be plotted. For example, `Model = c(1,1,1)` plots a model with three clusters and each cluster has 1 lag. Defaults to `Model = NULL`.
mar_heat	A numeric vector. Optional when using `show = "specific"` to specify the margins of the heatplot. Defaults to `mar_heat = c(2.5,2.5,2,1)`.
...	Pass additional arguments.

Details

Creates different plots showing either a fitted LCVAR model or fit indices for a specified set of LCVAR models.

Value

No return value, just plots figure.

Author(s)

Anja Ernst & Jonas Haslbeck

Examples

```
LCVAR_outExample <- LCVAR(Data = ExampleData,
                          yVars = 1:4, ID = 5, Beep = 6,
                          xContinuous = 7, xFactor = 8,
                          Clusters = 1:2, Lags = 1:2,
                          Center = FALSE, Cores = 2,
                          RndSeed = 3, Rand = 2,
                          it = 25)

plot(LCVAR_outExample, show = "GNL", Number_of_Lags = 1)
plot(LCVAR_outExample, show = "GNC", Number_of_Clusters = 2)
plot(LCVAR_outExample, show = "specific", Model = c(1, 1))
plot(LCVAR_outExample, show = "specific", Model = c(1, 1), labels = c("A", "B", "C","D"))
plot(LCVAR_outExample, show = "specificDiff", Model = c(1, 1))
```

---

print.ClusterVAR   *Prints description of ClusterVAR objects*

Description

Takes the output of the ClusterVAR object and prints a small overview of the fitted model(s).

Usage

```
## S3 method for class 'ClusterVAR'
print(x, ...)
```

Arguments

x	An output object of the LCVAR function
...	Pass additional arguments

Details

Prints an overview of the fitted model(s) in the console.

Value

No return value, just returns summary in console.

Author(s)

Anja Ernst & Jonas Haslbeck

---

print.ClusterVARCoef   *Overview of Parameters of given Model.*

Description

Overview of Parameters of given Model.

Usage

```
## S3 method for class 'ClusterVARCoef'
print(x, ...)
```

Arguments

x   An output object of the `coef.ClusterVAR` function.

...   Pass additional arguments.

Value

Prints an overview of the fitted model in the console.

Author(s)

Anja Ernst & Jonas Haslbeck

---

print.ClusterVARSummary
                    Print Summary of Models into Console.

Description

Print Summary of Models into Console.

Usage

```
## S3 method for class 'ClusterVARSummary'
print(x, ...)
```

Arguments

x	An output object of the `summary.ClusterVAR` function.
...	Pass additional arguments.

Value

Prints the summary of the fitted models in the console.

Author(s)

Anja Ernst & Jonas Haslbeck

---

print.PredictableObs   *Print Summary of predictable Observations into Console.*

Description

Print Summary of predictable Observations into Console.

Usage

```
    ## S3 method for class 'PredictableObs'
print(x, ...)
```

Arguments

x	An output object of the `numberPredictableObservations` function.
...	Pass additional arguments.

Value

Prints the summary of the number of predictable Observations in the console.

Author(s)

Anja Ernst & Jonas Haslbeck

---

summary.ClusterVAR   *Summary of ClusterVAR objects*

Description

Takes the output of the LCVAR function and creates a small summary of the fitted model(s).

Usage

```
## S3 method for class 'ClusterVAR'
summary(object,
  show = "BPC",
  TS_criterion = "SC",
  global_criterion = "BIC",
  Number_of_Clusters = NULL,
  Number_of_Lags = NULL,
  ...)
```

Arguments

object	An output object of the LCVAR function.
show	Indicate how models should be summarized, the possible choices are "BPC", "GNC" and "GNL". show = "BPC" compares models with different time lags for each number of clusters. If show = "BPC", for each number of clusters the model with the best lag is selected and displayed in the output. The best lag is selected through the time-series information criterion specified with the argument TS_criterion (see below). show = "GNC" shows all models with different lags for a given number of clusters, this number of clusters is specified through Number_of_Clusters (see below). show = "GNL" shows for each number of clusters the model where all lags are fixed to a given number, this number of lags is specified through Number_of_Lags (see below). Out of these models, the best model in terms of the number of clusters is selected by the information criterion selected with the argument global_criterion (see below). Defaults to show = "BPC".
TS_criterion	The information criterion to select the best model between models with a different number of lags but with the same number of clusters. The possible choices are "SC" and "HQ". Defaults to TS_criterion = "SC".
global_criterion	
	The information criterion to select the best model between models with different numbers of clusters but with the same number of lags. The possible choices are "BIC" and "ICL". Defaults to global_criterion = "BIC".
Number_of_Clusters	
	An integer. Specify the fixed number of clusters when using show = "GNC". Defaults to Number_of_Clusters = NULL.
Number_of_Lags	An integer. Specify the fixed number of lags when using show = "GNL". Defaults to Number_of_Lags = the lowest number of lags specified in object.
...	Pass additional arguments.

Value

`FunctionOutput` Is a data frame containing summaries of the fitted models.

Author(s)

Anja Ernst & Jonas Haslbeck

References

Hamilton, J. (1994), Time Series Analysis, Princeton University Press, Princeton.

Hannan, E. J. and B. G. Quinn (1979), The determination of the order of an autoregression, Journal of the Royal Statistical Society.

Lütkepohl, H. (2006), New Introduction to Multiple Time Series Analysis, Springer, New York.

Quinn, B. (1980), Order determination for a multivariate autoregression, Journal of the Royal Statistical Society.

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics.

See Also

`plot.ClusterVAR()`, `coef.ClusterVAR()`

Examples

```
LCVAR_outExample <- LCVAR(Data = ExampleData,
                          yVars = 1:4, ID = 5, Beep = 6,
                          xContinuous = 7, xFactor = 8,
                          Clusters = 1:2, Lags = 1:2,
                          Cores = 2, RndSeed = 3,
                          Rand = 2, it = 25)

summary(LCVAR_outExample)
summary(object = LCVAR_outExample, show = "GNL", Number_of_Lags = 1)
summary(object = LCVAR_outExample, show = "GNL", Number_of_Lags = 1, global_criterion = "ICL")
summary(object = LCVAR_outExample, show = "GNC", Number_of_Clusters = 2)
summary(object = LCVAR_outExample, show = "GNC", Number_of_Clusters = 2, TS_criterion = "HQ")
```

Index