

Package ‘gimme’

February 28, 2019

Title Group Iterative Multiple Model Estimation

Version 0.5-1

Date 2019-02-18

Author Stephanie Lane [aut, cre, trl],
Kathleen Gates [aut],
Zachary Fisher [aut],
Peter Molenaar [aut],
Michael Hallquist [ctb],
Hallie Pike [ctb],
Cara Arizmendi [ctb],
Teague Henry [ctb],
Kelly Duffy [ctb]

Maintainer KM Gates <gateskm@email.unc.edu>

Depends R (>= 3.0.2)

Imports lavaan (>= 0.5-19), igraph (>= 1.0-0), qgraph, data.tree,
MIIVsem, imputeTS

Description Automated identification and estimation of group- and individual-level relations in time series data from within a structural equation modeling framework.

License GPL-2

LazyData true

URL <https://github.com/GatesLab/gimme/>

BugReports <https://github.com/GatesLab/gimme/issues>

ByteCompile true

RoxygenNote 6.1.1

NeedsCompilation no

Suggests knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

Repository CRAN

Date/Publication 2019-02-28 18:00:03 UTC

R topics documented:

gimme-package	2
aggSEM	3
gimmeSEM	5
indSEM	9
ms.fit	12
simData	12
solution.tree	13
ts	13

Index	14
--------------	-----------

gimme-package	<i>Group iterative multiple model estimation</i>
---------------	--

Description

This package contains functions to identify group- and individual-level unified structural equation models.

Details

Researchers across varied domains gather multivariate data for each individual unit of study across multiple occasions of measurement. Generally referred to as time series (or in the social sciences, intensive longitudinal) data, examples include psychophysiological processes such as neuroimaging and heart rate variability, daily diary studies, and observational coding of social interactions among dyads. A primary goal for acquiring these data is to understand temporal processes. The `gimme` package contains several functions for use with these data. These functions include `gimmeSEM`, which provides both group- and individual-level results by looking across individuals for patterns of relations among variables. A function that provides group-level results, `aggSEM`, is included, as well as a function that provides individual-level results, `indSEM`. The major functions within the `gimme` package all require the user to specify the directory containing the data and a directory for output to be stored.

Author(s)

Stephanie Lane [aut, cre, trl],
 Kathleen Gates [aut],
 Zachary Fisher [aut],
 Peter Molenaar [aut],
 Michael Hallquist [ctb],
 Hallie Pike [ctb],
 Cara Arizmendi [ctb],
 Teague Henry [ctb], Kelly Duffy [ctb]
 Maintainer: KM Gates <gateskm@email.unc.edu>

aggSEM

*Group-level structural equation model search.***Description**

Concatenates all individual-level data files and fits a group model to the data.

Usage

```
aggSEM(data = "",
        out  = "",
        sep  = "",
        header = ,
        ar   = TRUE,
        plot = TRUE,
        paths = NULL,
        exogenous = NULL,
        conv_vars = NULL,
        conv_length = 16,
        conv_interval = 1,
        mult_vars = NULL,
        mean_center_mult = FALSE,
        standardize = FALSE)
```

Arguments

data	The path to the directory where the data files are located, or the name of the list containing each individual's time series. Each file or matrix must contain one matrix for each individual containing a T (time) by p (number of variables) matrix where the columns represent variables and the rows represent time. If in list form, each item in the list (i.e., matrix) must be named.
out	The path to the directory where the results will be stored (optional). If specified, a copy of output files will be replaced in directory. If directory at specified path does not exist, it will be created.
sep	The spacing of the data files when data are in a directory. "" indicates space-delimited, "/t" indicates tab-delimited, "," indicates comma delimited. Only necessary to specify if reading data in from physical directory.
header	Logical. Indicate TRUE for data files with a header. Only necessary to specify if reading data in from physical directory.
ar	Logical. If TRUE, begins search for group model with autoregressive (AR) paths open. Defaults to TRUE.
plot	Logical. If TRUE, figures depicting relations among variables of interest will automatically be created. For aggregate-level plot, red paths represent positive weights and blue paths represent negative weights. Dashed lines denote lagged relations (lag 1) and solid lines are contemporaneous (lag 0). Defaults to TRUE.

paths	Lavaan-style syntax containing paths with which to begin model estimation (optional). That is, $Y \sim X$ indicates that Y is regressed on X , or X predicts Y . If no header is used, then variables should be referred to with V followed (with no separation) by the column number. If a header is used, variables should be referred to using variable names. To reference lag variables, "lag" should be added to the end of the variable name with no separation. Defaults to NULL.
exogenous	Vector of variable names to be treated as exogenous (optional). That is, exogenous variable X can predict Y but cannot be predicted by Y . If no header is used, then variables should be referred to with V followed (with no separation) by the column number. If a header is used, variables should be referred to using variable names. Defaults to NULL.
conv_vars	Vector of variable names to be convolved via smoothed Finite Impulse Response (sFIR). Defaults to NULL.
conv_length	Expected response length in seconds. For functional MRI BOLD, 16 seconds (default) is typical for the hemodynamic response function.
conv_interval	Interval between data acquisition. Currently must be a constant. For fMRI studies, this is the repetition time. Defaults to 1.
mult_vars	Vector of variable names to be multiplied to explore bilinear/modulatory effects (optional). All multiplied variables will be treated as exogenous (X can predict Y but cannot be predicted by Y). Within the vector, multiplication of two variables should be indicated with an asterik (e.g. $V1 * V2$). If no header is used, variables should be referred to with V followed by the column number (with no separation). If a header is used, each variable should be referred to using variable names. If multiplication with the lag 1 of a variable is desired, the variable name should be followed by "lag" with no separation (e.g. $V1 * V2lag$). Note that if multiplied variables are desired, at least one variable in the dataset must be specified as exogenous. Defaults to NULL.
mean_center_mult	Logical. If TRUE, the variables indicated in <code>mult_vars</code> will be mean-centered before being multiplied together. Defaults to FALSE.
standardize	Logical. If TRUE, all variables will be standardized to have a mean of zero and a standard deviation of one. Defaults to FALSE. @details In main output directory: <ul style="list-style-type: none"> • allBetas Matrix. Contains estimates for each path in the aggregate-level model. The row variable is the outcome and the column variable is the predictor variable. • allStdErrors Matrix. Contains standard errors for each path in the aggregate-level model. The row variable is the outcome and the column variable is the predictor variable. • allPathEstimates Contains estimate, standard error, p-value, and z-value for each path for the concatenated data. • summaryFit Contains model fit information for the aggregate-level model. • summaryPathsPlot Contains aggregate-level plot. Red paths represent positive weights and blue paths represent negative weights.

Author(s)

Stephanie Lane

Examples

```
## Not run:  
exFit <- aggSEM(data = ts)  
  
## End(Not run)  
  
plot(exFit)
```

gimmeSEM

Group iterative multiple model estimation.

Description

This function identifies structural equation models for each individual that consist of both group-level and individual-level paths.

Usage

```
gimmeSEM(data      = "",  
          out       = "",  
          sep       = "",  
          header    = ,  
          ar        = TRUE,  
          plot      = TRUE,  
          subgroup  = FALSE,  
          sub_feature = "lag & contemp",  
          confirm_subgroup = NULL,  
          paths     = NULL,  
          exogenous = NULL,  
          conv_vars  = NULL,  
          conv_length = 16,  
          conv_interval = 1,  
          mult_vars  = NULL,  
          mean_center_mult = FALSE,  
          standardize = FALSE,  
          groupcutoff = .75,  
          subcutoff   = .5,  
          diagnos     = FALSE,  
          ms_allow    = FALSE,  
          ms_tol      = 1e-5,  
          lv_model    = NULL,  
          lv_scores   = "regression",  
          lv_estimator = "miiv",
```

```
lv_miiv_scaling = "first.indicator",
lv_final_estimator = "miiv")
```

Arguments

data	The path to the directory where the data files are located, or the name of the list containing each individual's time series. Each file or matrix must contain one matrix for each individual containing a T (time) by p (number of variables) matrix where the columns represent variables and the rows represent time. Individuals must have the same variables (p) but can have different lengths of observations (T).
out	The path to the directory where the results will be stored (optional). If specified, a copy of output files will be replaced in directory. If directory at specified path does not exist, it will be created.
sep	The spacing of the data files. "" indicates space-delimited, "/t" indicates tab-delimited, "," indicates comma delimited. Only necessary to specify if reading data in from physical directory.
header	Logical. Indicate TRUE for data files with a header. Only necessary to specify if reading data in from physical directory.
ar	Logical. If TRUE, begins search for group model with autoregressive (AR) paths freed for estimation. If ms_allow=TRUE, it is recommended to set ar=FALSE. Multiple solutions are unlikely to be found when ar=TRUE. Defaults to TRUE.
plot	Logical. If TRUE, graphs depicting relations among variables of interest will automatically be created. Solid lines represent contemporaneous relations (lag 0) and dashed lines reflect lagged relations (lag 1). For individual-level plots, red paths represent positive weights and blue paths represent negative weights. Width of paths corresponds to estimated path weight. For the group-level plot, black represents group-level paths, grey represents individual-level paths, and (if subgroup = TRUE) green represents subgroup-level paths. For the group-level plot, the width of the edge corresponds to the count. Defaults to TRUE.
subgroup	Logical. If TRUE, subgroups are generated based on similarities in model features using the walktrap.community function from the igraph package. When ms_allow=TRUE, subgroup should be set to FALSE. Defaults to FALSE.
sub_feature	Option to indicate feature(s) used to subgroup individuals. Defaults to "lag & contemp" for lagged and contemporaneous, which is the original method. Can use "lagged" or "contemp" to subgroup solely on features related to lagged and contemporaneous relations, respectively.
confirm_subgroup	Dataframe. Option only available when subgroup = TRUE. Dataframe should contain two columns. The first column should specify file labels (the name of the data files without file extension), and the second should contain integer values (beginning at 1) specifying the subgroup membership for each individual. function from the igraph package. Defaults to TRUE.
paths	lavaan-style syntax containing paths with which to begin model estimation (optional). That is, Y~X indicates that Y is regressed on X, or X predicts Y. If no header is used, then variables should be referred to with V followed (with no separation) by the column number. If a header is used, variables should be referred

to using variable names. To reference lag variables, "lag" should be added to the end of the variable name with no separation. Defaults to NULL. ##### Need to explain mult_vars too. #####

exogenous	Vector of variable names to be treated as exogenous (optional). That is, exogenous variable X can predict Y but cannot be predicted by Y. If no header is used, then variables should be referred to with V followed (with no separation) by the column number. If a header is used, variables should be referred to using variable names. Defaults to NULL.
conv_vars	Vector of variable names to be convolved via smoothed Finite Impulse Response (sFIR). Note, conv_vars are not automatically considered exogenous variables. To treat conv_vars as exogenous use the exogenous argument. Variables listed in conv_vars must be binary variables. If there is missing data in the endogenous variables their values will be imputed for the convolution operation only. Defaults to NULL. ### If there are multiple variables listed in conv_vars they are not used in the convolution of additional conv_vars.## You can't do lagged variables.
conv_length	Expected response length in seconds. For functional MRI BOLD, 16 seconds (default) is typical for the hemodynamic response function.
conv_interval	Interval between data acquisition. Currently must be a constant. For fMRI studies, this is the repetition time. Defaults to 1.
mult_vars	Vector of variable names to be multiplied to explore bilinear/modulatory effects (optional). All multiplied variables will be treated as exogenous (X can predict Y but cannot be predicted by Y). Within the vector, multiplication of two variables should be indicated with an asterik (e.g. V1*V2). If no header is used, variables should be referred to with V followed by the column number (with no separation). If a header is used, each variable should be referred to using variable names. If multiplication with the lag 1 of a variable is desired, the variable name should be followed by "lag" with no separation (e.g. V1*V2lag).
mean_center_mult	Logical. If TRUE, the variables indicated in mult_vars will be mean-centered before being multiplied together. Defaults to FALSE.
standardize	Logical. If TRUE, all variables will be standardized to have a mean of zero and a standard deviation of one. Defaults to FALSE.
groupcutoff	Cutoff value for group-level paths. Defaults to .75, indicating that a path must be significant across 75% of individuals to be included as a group-level path.
subcutoff	Cutoff value for subgroup- level paths. Defaults to .5, indicating that a path must be significant across at least 50% of the individuals in a subgroup to be considered a subgroup-level path.
diagnos	Logical.If TRUE provides internal output for diagnostic purposes. Defaults to FALSE.
ms_allow	Logical. If TRUE provides multiple solutions when more than one path has identical modification index values. When ms_allow=TRUE, it is recommended to set ar=FALSE. Multiple solutions are unlikely to be found when ar=TRUE. Additionally, subgroup should be set to FALSE.

<code>ms_tol</code>	Precision used when evaluating similarity of modification indices when <code>ms_allow = TRUE</code> . We recommend that <code>ms_tol</code> not be greater than the default, especially when <code>standardize=TRUE</code> . Defaults to <code>1e-5</code> .
<code>lv_model</code>	Invoke latent variable modeling by providing the measurement model syntax here. lavaan conventions are used for relating observed variables to factors. Defaults to <code>NULL</code>
<code>lv_scores</code>	Method used for estimating latent variable scores from parameters obtained from the factor analysis when <code>lv_model</code> is not <code>NULL</code> . Options are: "regression" (Default) or "bartlett".
<code>lv_estimator</code>	Estimator used for factor analysis. Options are "miiv" (default) or "pml" (pseudo-ML).
<code>lv_miiv_scaling</code>	Type of scaling indicator to use when "miiv" selected for <code>lv_estimator</code> . Options are "first.indicator" (Default; the first observed variable in the measurement equation is used), "group" (best one for the group), or "individual" (each individual has the best one for them according to R2).
<code>lv_final_estimator</code>	Estimator for final estimations. "miiv" (Default) or "pml" (pseudo-ML).

Details

In main output directory:

- **indivPathEstimates** Contains estimate, standard error, p-value, and z-value for each path for each individual. If `subgroup = TRUE` and subgroups are found, then a column is present containing the subgroup membership for each individual. Also contains the level at which each path was estimated: group, subgroup, or individual.
- **summaryFit** Contains model fit information for individual- level models. If subgroups are requested, this file also contains the subgroup membership for each individual.
- **summaryPathCountMatrix** Contains counts of total number of paths, both contemporaneous and lagged, estimated for the sample. The row variable is the outcome and the column variable is the predictor variable.
- **summaryPathCounts** Contains summary count information for paths identified at the group-, subgroup (if `subgroup = TRUE`), and individual-level.
- **summaryPathsPlot** Produced if `plot = TRUE`. Contains figure with group, subgroup (if `subgroup = TRUE`), and individual-level paths for the sample. Black paths are group-level, green paths are subgroup-level, and grey paths are individual-level, where the thickness of the line represents the count.

In subgroup output directory (if `subgroup = TRUE`):

- **subgroupkPathCounts** Contains counts of relations among lagged and contemporaneous variables for the *k*th subgroup.
- **subgroupkPlot** Contains plot of group, subgroup, and individual level paths for the *k*th subgroup. Black represents group-level paths, grey represents individual-level paths, and green represents subgroup-level paths.

Note: if a subgroup of size $n = 1$ is discovered, subgroup-level output is not produced.
 In individual output directory (where *id* represents the original file name for each individual):

- **idBetas** Contains individual-level estimates of each path for each individual.
- **idStdErrors** Contains individual-level standard errors for each path for each individual.
- **idPlot** Contains individual-level plots. Red paths represent positive weights and blue paths represent negative weights.

Author(s)

Stephanie Lane and Zachary Fisher

References

- Gates, K.M. & Molenaar, P.C.M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63, 310-319.
- Lane, S.T. & Gates, K.M. (2017). Automated selection of robust individual-level structural equation models for time series data. *Structural Equation Modeling*.
- Adriene M. Beltz & Peter C. M. Molenaar (2016) Dealing with Multiple Solutions in Structural Vector Autoregressive Models, *Multivariate Behavioral Research*, 51:2-3, 357-373.

Examples

```
## Not run:
paths <- 'V2 ~ V1
         V3 ~ V4lag'

fit <- gimmeSEM(data      = simData,
               out       = "C:/simData_out",
               subgroup  = TRUE,
               paths     = paths)

print(fit, mean = TRUE)
print(fit, subgroup = 1, mean = TRUE)
print(fit, file = "group_1_1", estimates = TRUE)
print(fit, subgroup = 2, fitMeasures = TRUE)
plot(fit, file = "group_1_1")

## End(Not run)
```

indSEM

Individual-level structural equation model search.

Description

This function identifies structural equation models for each individual. It does not utilize any shared information from the sample.

Usage

```

indSEM(data = NULL,
       out = NULL,
       sep = NULL,
       header = NULL,
       ar = TRUE,
       plot = TRUE,
       paths = NULL,
       exogenous = NULL,
       conv_vars = NULL,
       conv_length = 16,
       conv_interval = 1,
       mult_vars = NULL,
       mean_center_mult = FALSE,
       standardize = FALSE)

```

Arguments

data	The path to the directory where the data files are located, or the name of the list containing each individual's time series. Each file or matrix must contain one matrix for each individual containing a T (time) by p (number of variables) matrix where the columns represent variables and the rows represent time.
out	The path to the directory where the results will be stored (optional). If specified, a copy of output files will be replaced in directory. If directory at specified path does not exist, it will be created.
sep	The spacing of the data files. "" indicates space-delimited, "/t" indicates tab-delimited, "," indicates comma delimited. Only necessary to specify if reading data in from physical directory.
header	Logical. Indicate TRUE for data files with a header. Only necessary to specify if reading data in from physical directory.
ar	Logical. If TRUE, begins search for individual models with autoregressive (AR) paths open. Defaults to TRUE.
plot	Logical. If TRUE, graphs depicting relations among variables of interest will automatically be created. Defaults to TRUE. For individual-level plots, red paths represent positive weights and blue paths represent negative weights.
paths	lavaan-style syntax containing paths with which to begin model estimation. That is, Y~X indicates that Y is regressed on X, or X predicts Y. If no header is used, then variables should be referred to with V followed (with no separation) by the column number. If a header is used, variables should be referred to using variable names. To reference lag variables, "lag" should be added to the end of the variable name with no separation. Defaults to NULL.
exogenous	Vector of variable names to be treated as exogenous. That is, exogenous variable X can predict Y but cannot be predicted by Y. If no header is used, then variables should be referred to with V followed (with no separation) by the column number. If a header is used, variables should be referred to using variable names. Defaults to NULL.

conv_vars	Vector of variable names to be convolved via smoothed Finite Impulse Response (sFIR). Defaults to NULL.
conv_length	Expected response length in seconds. For functional MRI BOLD, 16 seconds (default) is typical for the hemodynamic response function.
conv_interval	Interval between data acquisition. Currently must be a constant. For fMRI studies, this is the repetition time. Defaults to 1.
mult_vars	Vector of variable names to be multiplied to explore bilinear/modulatory effects (optional). All multiplied variables will be treated as exogenous (X can predict Y but cannot be predicted by Y). Within the vector, multiplication of two variables should be indicated with an asterik (e.g. V1*V2). If no header is used, variables should be referred to with V followed by the column number (with no separation). If a header is used, each variable should be referred to using variable names. If multiplication with the lag 1 of a variable is desired, the variable name should be followed by "lag" with no separation (e.g. V1*V2lag). Note that if multiplied variables are desired, at least one variable in the dataset must be specified as exogenous. Defaults to NULL.
mean_center_mult	Logical. If TRUE, the variables indicated in mult_vars will be mean-centered before being multiplied together. Defaults to FALSE.
standardize	Logical. If TRUE, all variables will be standardized to have a mean of zero and a standard deviation of one. Defaults to FALSE.

Details

In main output directory:

- **indivPathEstimates** Contains estimate, standard error, p-value, and z-value for each path for each individual
- **summaryFit** Contains model fit information for individual-level models.
- **summaryPathCountMatrix** Contains counts of total number of paths, both contemporaneous and lagged, estimated for the sample. The row variable is the outcome and the column variable is the predictor variable.
- **summaryPathCounts** Contains summary count information for paths identified at the individual-level.
- **summaryPathsPlot** Contains counts of total number of paths, both contemporaneous and lagged, estimated for the sample. The row variable is the outcome and the column variable is the predictor variable.

In individual output directory (where *id* represents the original file name for each individual):

- **idBetas** Contains individual-level estimates of each path for each individual.
- **idStdErrors** Contains individual-level standard errors for each path for each individual.
- **idPlot** Contains individual-level plots. Red paths represent positive weights and blue paths represent negative weights.

Author(s)

Stephanie Lane

Examples

```
## Not run:
fit <- indSEM(data = "C:/data100",
              out  = "C:/data100_indSEM_out",
              sep  = ",",
              header = FALSE)
print(fit, file = "group1.1", estimates = TRUE)
plot(fit, file = "group1.1")

## End(Not run)
```

ms.fit	<i>Fitted gimme object with multiple solutions</i>
--------	--

Description

This object contains a fitted gimme object where multiple solutions gimme was used. The simulated data had 25 individuals, each with 100 time points.

Usage

```
ms.fit
```

Format

A fitted gimme object, where multiple solutions gimme was used.

simData	<i>Large example, heterogeneous data, group, subgroup, and individual level effects.</i>
---------	--

Description

This object contains a list of simulated time series data for twenty-five individuals with 200 time points and 10 variables, or regions of interest.

Usage

```
simData
```

Format

A list of data frames with 25 individuals, who each have 200 observations on 10 variables.

solution.tree	<i>Solution trees for multiple solutions gimme.</i>
---------------	---

Description

This function allows for the exploration of divergences in multiple solutions gimme for both the group and individuals.

Usage

```
solution.tree(x,
              level = c("group", "individual"),
              cols = NULL,
              ids = "all",
              plot.tree = FALSE)
```

Arguments

x	A fitted gimme object.
level	A character vector indicating what levels of the solution tree you would like returned. Options are "group", "individual", or c("group", "individual"). Defaults to c("group", "individual").
cols	A character vector indicating additional information to include in tree plot. Options include "stage", "pruned", "rmsea", "nnfi", "cfi", "srmr", "grp_sol", "bic", "aic", "modularity." Defaults to NULL.
ids	A character vector indicating the names of subjects to print. Defaults to "all."
plot.tree	Logical. If TRUE, plot of tree is produced. Defaults to FALSE.

Details

solution.tree

ts	<i>Small example, heterogeneous data, group and individual level effects</i>
----	--

Description

This object contains a list of simulated time series data for five individuals with 50 time points and 3 variables, or regions of interest.

Usage

```
ts
```

Format

A list of data frames with 5 individuals, who each have 50 observations on 3 variables.

Index

*Topic **datasets**

ms.fit, [12](#)

simData, [12](#)

ts, [13](#)

*Topic **gimmeSEM**

gimmeSEM, [5](#)

*Topic **gimme**

gimme-package, [2](#)

*Topic **indSEM**

indSEM, [9](#)

aggSEM, [2, 3](#)

gimme (gimmeSEM), [5](#)

gimme-package, [2](#)

gimmeSEM, [2, 5](#)

indSEM, [2, 9](#)

ms.fit, [12](#)

simData, [12](#)

solution.tree, [13](#)

ts, [13](#)