

Package ‘gsynth’

October 16, 2017

Type Package

Title Generalized Synthetic Control Method

Version 1.0.6

Date 2017-10-14

Author Yiqing Xu, Licheng Liu

Maintainer Yiqing Xu <yiqingxu@ucsd.edu>

Description Provides causal inference with interactive fixed-effect models. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effects model that incorporates unit-specific intercepts interacted with time-varying coefficients. This method generalizes the synthetic control method to the case of multiple treated units and variable treatment periods, and improves efficiency and interpretability. This version supports unbalanced panels. Reference: Yiqing Xu (2017) <doi:10.1017/pan.2016.2>.

URL http://yiqingxu.org/software/gsynth/gsynth_examples.html

NeedsCompilation yes

License GPL-2

Imports Rcpp (>= 0.12.3), ggplot2 (>= 2.1.0), GGally (>= 1.0.1), doParallel (>= 1.0.10), foreach (>= 1.4.3), abind (>= 1.4-0), mvtnorm (>= 1.0-6)

SystemRequirements A C++11 compiler.

Depends R (>= 2.10)

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.0.1

Repository CRAN

Date/Publication 2017-10-16 17:57:09 UTC

R topics documented:

gsynth-package	2
gsynth	3

gsynth-internal	6
interFE	7
plot.gsynth	9
print.gsynth	10
print.interFE	11
simdata	12
turnout	12
Index	13

gsynth-package	<i>Generalized Synthetic Control Method</i>
----------------	---

Description

Implements the generalized synthetic control method based on interactive fixed effect models.

Details

Implements the generalized synthetic control method. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effects model that incorporates unit-specific intercepts interacted with time-varying coefficients.

See [gsynth](#) for details.

Author(s)

Yiqing Xu <yiqingxu@ucsd.edu>, University of California, San Diego

Licheng Liu <liulch.16@sem.tsinghua.edu.cn>, Tsinghua University

References

Yiqing Xu. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76. Available at: <https://doi.org/10.1017/pan.2016.2>.

For more details, see http://yiqingxu.org/software/gsynth/gsynth_examples.html.

 gsynth

Generalized Synthetic Control Method

Description

Implements the generalized synthetic control method based on interactive fixed effect models.

Usage

```
gsynth(formula=NULL, data, Y, D, X=NULL, na.rm = FALSE, index,
       force = "unit", r = 0, CV = TRUE, EM = FALSE, se = FALSE, nboots = 200,
       inference = "nonparametric", cov.ar = 1, parallel = FALSE,
       cores = NULL, tol = 0.001, seed = NULL, min.T0 = 5, normalize = FALSE)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted.
data	a data frame (must be with a dichotomous treatment but balanced is not required).
Y	outcome.
D	treatment.
X	time-varying covariates.
na.rm	a logical flag indicating whether to list-wise delete missing data. The algorithm will report an error if missing data exist.
index	a two-element string vector specifying the unit (group) and time indicators. Must be of length 2.
force	a string indicating whether unit or time fixed effects will be imposed. Must be one of the following, "none", "unit", "time", or "two-way". The default is "unit".
r	an integer specifying the number of factors. If CV = TRUE, the cross validation procedure will select the optimal number of factors from r to 5.
CV	a logical flag indicating whether cross-validation will be performed to select the optimal number of factors. If r is not specified, the procedure will search through r = 0 to 5.
EM	a logical flag indicating whether an Expectation Maximization algorithm will be used (Gobillon and Magnac 2016).
se	a logical flag indicating whether uncertainty estimates will be produced.
nboots	an integer specifying the number of bootstrap runs. Ignored if se = FALSE.
inference	a string specifying which type of inferential method will be used, either "parametric" or "non-parametric". "parametric" is recommended when the number of treated units is small.
cov.ar	an integer specifying order of the auto regression process that the residuals follow. Used for parametric bootstrap procedure when data is in the form of unbalanced panel.

parallel	a logical flag indicating whether parallel computing will be used in bootstrapping and/or cross-validation. Ignored if <code>se = FALSE</code> .
cores	an integer indicating the number of cores to be used in parallel computing. If not specified, the algorithm will use the maximum number of logical cores of your computer (warning: this could prevent you from multi-tasking on your computer).
tol	a positive number indicating the tolerance level.
seed	an integer that sets the seed in random number generation. Ignored if <code>se = FALSE</code> and <code>r</code> is specified.
min.T0	an integer specifying the minimum value of pre-treatment periods. Treated units with pre-treatment periods less than that will be removed automatically.
normalize	a logic flag indicating whether to scale outcome and covariates. Useful for accelerating computing speed when magnitude of data is large. The default is <code>normalize=FALSE</code> .

Details

gsynth implements the generalized synthetic control method. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effects model that incorporates unit-specific intercepts interacted with time-varying coefficients. It generalizes the synthetic control method to the case of multiple treated units and variable treatment periods, and improves efficiency and interpretability. It allows the treatment to be correlated with unobserved unit and time heterogeneities under reasonable modelling assumptions. With a built-in cross-validation procedure, it avoids specification searches and thus is easy to implement. Data must be with a dichotomous treatment.

Value

Y.dat	a matrix storing data of the outcome variable.
Y	name of the outcome variable.
D	name of the treatment variable.
X	name of the time-varying control variables.
index	name of the unit and time indicators.
id	a vector of unit IDs.
time	a vector of time periods.
obs.missing	a matrix of indicators of observations.
id.tr	a vector of IDs for the treatment units.
id.co	a vector of IDs for the control units.
Y.tr	data of the treated unit outcome.
Y.ct	predicted counterfactuals for the treated units.
Y.co	data of the control unit outcome.
eff	difference between actual outcome and predicted $Y(0)$.
Y.bar	average values of Y.tr, Y.ct, and Y.co over time.

att	average treatment effect on the treated over time (it is averaged based on the timing of the treatment if it is different for each unit).
att.avg	average treatment effect on the treated.
force	user specified force option.
DID	TRUE if the timing of the treatment is the same.
T	the number of time periods.
N	the total number of units.
p	the number of time-varying observables.
Ntr	the number of treated units.
Nco	the number of control units.
T0	a vector that stores the timing of the treatment for balanced panel data or pre-treatment periods of each treated unit for unbalanced panel data.
T0.ub	a vector that stores the timing of the treatment for unbalanced panel data.
tr	a vector indicating treatment status for each unit.
pre	a matrix indicating the pre-treatment/non-treatment status.
post	a matrix indicating the post-treatment status.
r.cv	the number of factors included in the model – either supplied by users or automatically chosen via cross-validation.
res.co	residuals of the control group units.
beta	coefficients of time-varying observables from the interactive fixed effect model.
sigma2	the mean squared error of interactive fixed effect model.
IC	the information criterion.
est.co	result of the interactive fixed effect model based on the control group data. An interFE object.
eff.cnt	difference between actual outcome and predicted $Y(0)$; rearranged based on the timing of the treatment.
Y.tr.cnt	data of the treated unit outcome, rearranged based on the timing of the treatment.
Y.ct.cnt	data of the predicted $Y(0)$, rearranged based on the timing of the treatment.
MSPE	mean squared prediction error of the cross-validated model.
CV.out	result of the cross-validation procedure.
niter	the number of iterations in the estimation of the interactive fixed effect model.
factor	estimated time-varying factors.
lambda.co	estimated loadings for the control group.
lambda.tr	estimated loadings for the treatment group.
mu	estimated ground mean.
xi	estimated time fixed effects.
alpha.tr	estimated unit fixed effects for the treated units.
alpha.co	estimated unit fixed effects for the control units.
obs.missing	a matrix storing status of each unit at each time point. 0 for missing, 1 for control group units, 2 for treat group units at pre-treatment period, 3 for treat group units at post-treatment period, and 4 for removed treated group units. Useful for unbalanced panel data.

Author(s)

Yiqing Xu <yiqingxu@ucsd.edu>

Licheng Liu <liulch.16@sem.tsinghua.edu.cn>

References

Laurent Gobillon and Thierry Magnac, 2016. "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls." *The Review of Economics and Statistics*, July 2016, Vol. 98, No. 3, pp. 535–551.

Yiqing Xu. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76. Available at: <https://doi.org/10.1017/pan.2016.2>.

For more details, see http://yiqingxu.org/software/gsynth/gsynth_examples.html.

See Also

[plot.gsynth](#) and [print.gsynth](#)

Examples

```
library(gsynth)
data(gsynth)
out <- gsynth(Y ~ D + X1 + X2, data = simdata,
             index = c("id", "time"), force = "two-way",
             CV = TRUE, r = c(0, 5), se = FALSE)
```

gsynth-internal

Internal Gsynth Functions

Description

Internal Gsynth functions

Details

These are not to be called by the user (or in some cases are just waiting for proper documentation to be written :).

interFE *Interactive Fixed Effects Models*

Description

Estimating interactive fixed effect models.

Usage

```
interFE(formula = NULL, data, Y, X, index, r = 0, force = "none",
        se = TRUE, nboots = 500, seed = NULL, normalize = FALSE)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted.
data	a data frame (must be with a dichotomous treatment but balanced is not required).
Y	outcome.
X	time-varying covariates.
index	a two-element string vector specifying the unit (group) and time indicators. Must be of length 2.
r	an integer specifying the number of factors.
force	a string indicating whether unit or time fixed effects will be imposed. Must be one of the following, "none", "unit", "time", or "two-way". The default is "unit".
se	a logical flag indicating whether uncertainty estimates will be produced via bootstrapping.
nboots	an integer specifying the number of bootstrap runs. Ignored if se = FALSE.
seed	an integer that sets the seed in random number generation. Ignored if se = FALSE and r is specified.
normalize	a logic flag indicating whether to scale outcome and covariates. Useful for accelerating computing speed when magnitude of data is large. The default is normalize=FALSE.

Details

interFE estimates interactive fixed effect models proposed by Bai (2009).

Value

beta	estimated coefficients.
mu	estimated grand mean.
factor	estimated factors.
lambda	estimated factor loadings.

VNT	a diagonal matrix that consists of the r eigenvalues.
niter	the number of iteration before convergence.
alpha	estimated unit fixed effect (if force is "unit" or "two-way").
xi	estimated time fixed effect (if force is "time" or "two-way").
residuals	residuals of the estimated interactive fixed effect model.
sigma2	mean squared error of the residuals.
IC	the information criterion.
ValidX	a bool value specifying whether there are valid covariates.
dat.Y	a matrix storing data of the outcome variable.
dat.X	an array storing data of the independent variables.
Y	Name of the outcome variable.
D	Name of the treatment variable.
X	Name of the time-varying control variables.
index	Name of the unit and time indicators.
est.table	a table of the estimation results.
est.boot	a matrix storing results from bootstraps.

Author(s)

Yiqing Xu <yiqingxu@ucsd.edu>

Licheng Liu <liulch.16@sem.tsinghua.edu.cn>

References

Jushan Bai. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77:1229–1279.

See Also

[print.interFE](#) and [gsynth](#)

Examples

```
library(gsynth)
data(gsynth)
d <- simdata[-(1:150),] # remove the treated units
out <- interFE(Y ~ X1 + X2, data = d, index=c("id","time"),
              r = 2, force = "two-way", nboots = 50)
```


plot.gsynth

*Plotting***Description**

Visualizes estimation results of the generalized synthetic control method.

Usage

```
## S3 method for class 'gsynth'
plot(x, type = "gap", xlim = NULL, ylim = NULL,
      xlab = NULL, ylab = NULL, legendOff = FALSE, raw = "band",
      main = NULL, nfactors = NULL, id = NULL, axis.adjust = FALSE, ...)
```

Arguments

x	a gsynth object.
type	a string that specifies the type of the plot. Must be one of the following: "gap" (plotting the average treatment effect on the treated; "raw" (plotting the raw data); "counterfactual" (plotting predicted Y(0)'s); "factors" (plotting estimated factors); "loadings" (plotting the distribution of estimated factor loadings); "missing" (plotting status of each unit at each time point).
xlim	a two-element numeric vector specifying the range of x-axis. When class of time variable is string, must specify not original value but a counting number e.g. xlim=c(1,30).
ylim	a two-element numeric vector specifying the range of y-axis.
xlab	a string indicating the label of the x-axis.
ylab	a string indicating the label of the y-axis.
legendOff	a logical flag controlling whether to show the legend.
raw	a string indicating whether or how raw data for the outcome variable will be shown in the "counterfactual" plot. Ignored if type is not "counterfactual". Must be one of the following: "none" (not showing the raw data); "band" (showing the middle 90 percentiles of the raw data); and "all" (showing the raw data as they are).
main	a string that controls the title of the plot. If not supplied, no title will be shown.
nfactors	a positive integer that specifies the number of factors to be shown. The maximum number is 4. Ignored if type is not "factors"
id	a unit identifier of which the predicted counterfactual or the difference between actual and predicted counterfactual is to be shown. It can also be a vector specifying units to be plotted if type=="missing" when data magnitude is large. Ignored if type is none of "missing", "counterfactual", "gap".
axis.adjust	a logic flag indicating whether to adjust labels on x-axis. Useful when class of time variable is string and data magnitude is large.
...	other argv.

Details

plot.gsynth visualizes the raw data used by, or estimation results obtained from, the generalized synthetic control method.

Author(s)

Yiqing Xu <yiqingxu@ucsd.edu>

Licheng Liu <liulch.16@sem.tsinghua.edu.cn>

References

Yiqing Xu. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76. Available at: <https://doi.org/10.1017/pan.2016.2>.

See http://yiqingxu.org/software/gsynth/gsynth_examples.html for more detailed information.

See Also

[gsynth](#) and [print.gsynth](#)

print.gsynth

Print Results

Description

Print results of the generalized synthetic control method.

Usage

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

Arguments

x a [gsynth](#) object.
... other argv.

Author(s)

Yiqing Xu <yiqingxu@ucsd.edu>

Licheng Liu <liulch.16@sem.tsinghua.edu.cn>

References

Yiqing Xu. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76. Available at: <https://doi.org/10.1017/pan.2016.2>.

For more details, see http://yiqingxu.org/software/gsynth/gsynth_examples.html.

See Also

[gsynth](#) and [plot.gsynth](#)

print.interFE

Print Results

Description

Print results of interactive fixed effects estimation.

Usage

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

Arguments

x an [interFE](#) object.
... other argv.

Author(s)

Yiqing Xu <yiqingxu@ucsd.edu>

Licheng Liu <liulch.16@sem.tsinghua.edu.cn>

References

Jushan Bai. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77:1229–1279.

See Also

[interFE](#) and [gsynth](#)

simdata	<i>simdata</i>
---------	----------------

Description

A simulated dataset.

Format

dataframe

References

Yiqing Xu. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76. Available at: <https://doi.org/10.1017/pan.2016.2>.

For more details, see http://yiqingxu.org/software/gsynth/gsynth_examples.html.

turnout	<i>turnout</i>
---------	----------------

Description

State-level voter turnout data.

Format

dataframe

References

Melanie Jean Springer. 2014. *How the States Shaped the Nation: American Electoral Institutions and Voter Turnout, 1920-2000*. University of Chicago Press.

Yiqing Xu. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis*, Vol. 25, Iss. 1, January 2017, pp. 57-76. Available at: <https://doi.org/10.1017/pan.2016.2>.

For more details, see http://yiqingxu.org/software/gsynth/gsynth_examples.html.

Index

*Topic **datasets**

simdata, [12](#)

turnout, [12](#)

*Topic **ts**

gsynth-internal, [6](#)

_gsynth_XXinv (gsynth-internal), [6](#)

_gsynth_beta_iter (gsynth-internal), [6](#)

_gsynth_beta_iter_ub (gsynth-internal),
[6](#)

_gsynth_data_ub_adj (gsynth-internal), [6](#)

_gsynth_inter_fe (gsynth-internal), [6](#)

_gsynth_inter_fe_ub (gsynth-internal), [6](#)

_gsynth_panel_beta (gsynth-internal), [6](#)

_gsynth_panel_est (gsynth-internal), [6](#)

_gsynth_panel_factor (gsynth-internal),
[6](#)

_gsynth_panel_factor_ub
(gsynth-internal), [6](#)

beta_iter (gsynth-internal), [6](#)

beta_iter_ub (gsynth-internal), [6](#)

ct.adjstut (gsynth-internal), [6](#)

data_ub_adj (gsynth-internal), [6](#)

gsynth, [2](#), [3](#), [8–11](#)

gsynth-internal, [6](#)

gsynth-package, [2](#)

gsynth.default (gsynth-internal), [6](#)

gsynth.formula (gsynth-internal), [6](#)

inter_fe (gsynth-internal), [6](#)

inter_fe_ub (gsynth-internal), [6](#)

interFE, [5](#), [7](#), [11](#)

interFE.default (gsynth-internal), [6](#)

interFE.formula (gsynth-internal), [6](#)

panel_beta (gsynth-internal), [6](#)

panel_est (gsynth-internal), [6](#)

panel_factor (gsynth-internal), [6](#)

panel_factor_ub (gsynth-internal), [6](#)

plot.gsynth, [6](#), [9](#), [11](#)

print.gsynth, [6](#), [10](#), [10](#)

print.interFE, [8](#), [11](#)

remove_na (gsynth-internal), [6](#)

res.vcov (gsynth-internal), [6](#)

simdata, [12](#)

synth.boot (gsynth-internal), [6](#)

synth.core (gsynth-internal), [6](#)

synth.em (gsynth-internal), [6](#)

turnout, [12](#)

XXinv (gsynth-internal), [6](#)