Package: cdid (via r-universe)

January 9, 2025

Type Package

Title The Chained Difference-in-Differences

Version 0.1.0

URL https://github.com/joelcuerrier/cdid

Maintainer David Benatia <david.benatia@hec.ca>

Description Extends the 'did' package to improve efficiency and handling of unbalanced panel data. Bellego, Benatia, and Dortet-Bernadet (2024), ``The Chained Difference-in-Differences'', Journal of Econometrics, <doi:10.1016/j.jeconom.2024.105783>.

License GPL-2

Encoding UTF-8

LazyData true

Imports Matrix, purrr, MASS, did, BMisc, methods, stats, utils

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

BugReports https://github.com/joelcuerrier/cdid/issues

VignetteBuilder knitr

Suggests rmarkdown, knitr, openxlsx, data.table, jsonlite, stargazer, devtools,

Config/pak/sysreqs cmake make libicu-dev

Repository https://joelcuerrier.r-universe.dev

RemoteUrl https://github.com/joelcuerrier/cdid

RemoteRef HEAD

RemoteSha 3b1b15bb361acaf8325f4c90d4500797070e2f3b

Contents

att_gt_cdid	2
DIDparams	5
fonction_simu_attrition	8
gg	9
gmm_compute_delta_att	10
gmm_convert_delta_to_att	10
gmm_convert_result	11
MP	11
pre_process_cdid	12
print.MP	15
process_attgt_gmm	16
summary.MP	16
-	
	18

Index

att_gt_cdid

att_gt_cdid

Description

att_gt_cdid computes average treatment effects. Our estimator accommodates (1) multiple time periods, (2) variation in treatment timing, (3) treatment effect heterogeneity, and (4) general missing data patterns. For more details on the methodology, see: Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

```
att_gt_cdid(
  yname,
  tname,
  idname = NULL,
  gname,
  xformla = NULL,
  data,
  panel = TRUE,
  allow_unbalanced_panel = TRUE,
  control_group,
  anticipation = 0,
  weightsname = NULL,
  alp = 0.05,
  bstrap = TRUE,
  cband = TRUE,
  biters = 1000,
  clustervars = NULL,
  est_method = "2-step",
  base_period = "varying",
```

```
print_details = FALSE,
pl = FALSE,
cores = 1
)
```

yname	The name of the outcome variable
tname	The name of the column containing the time periods
idname	The individual (cross-sectional unit) id name
gname	The name of the variable in data that contains the first period when a particular observation is treated. This should be a positive number for all observations in treated groups. It defines which "group" a unit belongs to. It should be 0 for units in the untreated group.
xformla	A formula for the covariates to include in the model. It should be of the form ~ X1 + X2. Default is NULL which is equivalent to xformla=~1. This is used to create a matrix of covariates which is then passed to the 2x2 DID estimator chosen in est_method. X's are assumed fixed across the time dimension in this version. Use different columns Xt, Xt+1 if time-varying covariates are needed.
data	The name of the data.frame that contains the data
panel	(Not used) This is not used as balanced and unbalanced panel data is treated similarly.
allow_unbalanc	ed_panel
	(Not used) This is not used as balanced and unbalanced panel data is treated similarly.
control_group	Which units to use the control group. The default is "nevertreated" which sets the control group to be the group of units that never participate in the treatment. This group does not change across groups or time periods. The other option is to set group="notyettreated". In this case, the control group is set to the group of units that have not yet participated in the treatment in that time period. This includes all never treated units, but it includes additional units that eventually participate in the treatment, but have not participated yet.
anticipation	(Not used) The number of time periods before participating in the treatment where units can anticipate participating in the treatment and therefore it can affect their untreated potential outcomes
weightsname	The name of the column containing weights. If not set, all observations have same weight.
alp	the significance level, default is 0.05
bstrap	Boolean for whether or not to compute standard errors using the multiplier boot- strap. If standard errors are clustered, then one must set bstrap=TRUE. Default is TRUE (in addition, cband is also by default TRUE indicating that uniform confi- dence bands will be returned. If bstrap is FALSE, then analytical standard errors are reported.
cband	Boolean for whether or not to compute a uniform confidence band that covers all of the group-time average treatment effects with fixed probability 1-alp. In

order to compute uniform confidence bands, bstrap must also be set to TRUE. The default is TRUE.

- biters The number of bootstrap iterations to use. The default is 1000, and this is only applicable if bstrap=TRUE.
- clustervars A vector of variables names to cluster on. At most, there can be two variables (otherwise will throw an error) and one of these must be the same as idname which allows for clustering at the individual level. By default, we cluster at individual level (when bstrap=TRUE).
- est_method the method to compute group-time average treatment effects. At the moment, one can only use the IPW estimator with either "2-step" or "Identity" weighting matrix to aggregate Delta ATT into ATT. include "ipw" for inverse probability weighting and "reg" for first step regression estimators.
- base_period (Not used) The cdid package only uses the g-1 base period for the moment. Whether to use a "varying" base period or a "universal" base period. Either choice results in the same post-treatment estimates of ATT(g,t)'s. In pre-treatment periods, using a varying base period amounts to computing a pseudo-ATT in each treatment period by comparing the change in outcomes for a particular group relative to its comparison group in the pre-treatment periods (i.e., in pretreatment periods this setting computes changes from period t-1 to period t, but repeatedly changes the value of t)

A universal base period fixes the base period to always be (g-anticipation-1). This does not compute pseudo-ATT(g,t)'s in pre-treatment periods, but rather reports average changes in outcomes from period t to (g-anticipation-1) for a particular group relative to its comparison group. This is analogous to what is often reported in event study regressions.

Using a varying base period results in an estimate of ATT(g,t) being reported in the period immediately before treatment. Using a universal base period normalizes the estimate in the period right before treatment (or earlier when the user allows for anticipation) to be equal to 0, but one extra estimate in an earlier period.

print_detailsWhether or not to show details/progress of computations. Default is FALSE.plWhether or not to use parallel processingcoresThe number of cores to use for parallel processing

Value

an MP object containing all the results for group-time average treatment effects

References

Bellego, Benatia, and Dortet-Bernadet (2024) \"The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

DIDparams

Description

Creates a DIDparams object to hold parameters for difference-in-differences analysis, including data structure details and user-specified options. This object is designed to streamline parameter passing across functions in the cdid package.#'

Usage

```
DIDparams(
 yname,
  tname,
  idname = NULL,
  gname,
 xformla = NULL,
 data,
  control_group,
  anticipation = 0,
 weightsname = NULL,
  alp = 0.05,
 bstrap = TRUE,
  biters = 1000,
  clustervars = NULL,
  cband = TRUE,
  print_details = TRUE,
 pl = FALSE,
  cores = 1,
  est_method = "chained",
 base_period = "varying",
 panel = TRUE,
  true_repeated_cross_sections,
 n = NULL,
  nG = NULL,
 nT = NULL,
  tlist = NULL,
 glist = NULL,
  call = NULL
)
```

yname	The name of the outcome variable
tname	The name of the column containing the time periods
idname	The individual (cross-sectional unit) id name

gname	The name of the variable in data that contains the first period when a particular observation is treated. This should be a positive number for all observations in treated groups. It defines which "group" a unit belongs to. It should be 0 for units in the untreated group.
xformla	A formula for the covariates to include in the model. It should be of the form $\sim X1 + X2$. Default is NULL which is equivalent to xformla= ~1 . This is used to create a matrix of covariates which is then passed to the 2x2 DID estimator chosen in est_method. X's are assumed fixed across the time dimension in this version. Use different columns Xt, Xt+1 if time-varying covariates are needed.
data	The name of the data.frame that contains the data
control_group	Which units to use the control group. The default is "nevertreated" which sets the control group to be the group of units that never participate in the treatment. This group does not change across groups or time periods. The other option is to set group="notyettreated". In this case, the control group is set to the group of units that have not yet participated in the treatment in that time period. This includes all never treated units, but it includes additional units that eventually participate in the treatment, but have not participated yet.
anticipation	(Not used) The number of time periods before participating in the treatment where units can anticipate participating in the treatment and therefore it can affect their untreated potential outcomes
weightsname	The name of the column containing the sampling weights. If not set, all observations have same weight.
alp	the significance level, default is 0.05
bstrap	Boolean for whether or not to compute standard errors using the multiplier boot- strap. If standard errors are clustered, then one must set bstrap=TRUE. Default is TRUE (in addition, cband is also by default TRUE indicating that uniform confi- dence bands will be returned. If bstrap is FALSE, then analytical standard errors are reported.
biters	The number of bootstrap iterations to use. The default is 1000, and this is only applicable if bstrap=TRUE.
clustervars	A vector of variables names to cluster on. At most, there can be two variables (otherwise will throw an error) and one of these must be the same as idname which allows for clustering at the individual level. By default, we cluster at individual level (when bstrap=TRUE).
cband	Boolean for whether or not to compute a uniform confidence band that covers all of the group-time average treatment effects with fixed probability 1-alp. In order to compute uniform confidence bands, bstrap must also be set to TRUE. The default is TRUE.
print_details	Whether or not to show details/progress of computations. Default is FALSE.
pl	Whether or not to use parallel processing
cores	The number of cores to use for parallel processing
est_method	the method to compute group-time average treatment effects. At the moment, one can only use the IPW estimator with either "2-step" or "Identity" weighting matrix to aggregate Delta ATT into ATT. include "ipw" for inverse probability weighting and "reg" for first step regression estimators.

base_period	(Not used) The cdid package only uses the g-1 base period for the moment. Whether to use a "varying" base period or a "universal" base period. Either choice results in the same post-treatment estimates of $ATT(g,t)$'s. In pre-treatment periods, using a varying base period amounts to computing a pseudo-ATT in each treatment period by comparing the change in outcomes for a particular group relative to its comparison group in the pre-treatment periods (i.e., in pre- treatment periods this setting computes changes from period t-1 to period t, but repeatedly changes the value of t) A universal base period fixes the base period to always be (α -anticipation-1)
	This does not compute pseudo-ATT(g,t)'s in pre-treatment periods, but rather reports average changes in outcomes from period t to (g-anticipation-1) for a particular group relative to its comparison group. This is analogous to what is often reported in event study regressions.
	Using a varying base period results in an estimate of ATT(g,t) being reported in the period immediately before treatment. Using a universal base period nor- malizes the estimate in the period right before treatment (or earlier when the user allows for anticipation) to be equal to 0, but one extra estimate in an earlier period.
panel	(Not used) This is not used as balanced and unbalanced panel data is treated similarly.
true_repeated_	_cross_sections
	Whether or not the data really is repeated cross sections. (We include this be- cause unbalanced panel code runs through the repeated cross sections code)
n	The number of observations. This is equal to the number of units (which may be different from the number of rows in a panel dataset).
nG	The number of groups
nT	The number of time periods
tlist	a vector containing each time period
glist	a vector containing each group
call	(Not used) a call control var

Value

A DIDparams object, which is a list containing the following elements:

- yname: The name of the outcome variable.
- tname: The name of the time variable.
- idname: The name of the unit identifier variable (if applicable).
- gname: The name of the group variable (e.g., treatment group).
- xformla: A formula specifying covariates for the model.
- data: The dataset used for analysis.
- control_group: The type of control group (e.g., "never treated" or "not yet treated").
- anticipation: The number of periods of anticipation before treatment.
- weightsname: The name of the variable containing sampling weights (if applicable).

- alp: The significance level (default is 0.05).
- bstrap: Logical. Indicates whether bootstrap is used for standard errors.
- biters: The number of bootstrap iterations (if bootstrap is enabled).
- clustervars: Variables used for clustering standard errors.
- cband: Logical. Indicates whether simultaneous confidence bands are computed.
- print_details: Logical. Indicates whether detailed results should be printed.
- pl: Logical. Parallelization flag for computations.
- cores: The number of cores to use for parallelization (if enabled).
- est_method: The estimation method (e.g., "chained").
- base_period: The base period used for comparison (e.g., "varying").
- panel: Logical. Indicates whether the data is a panel dataset.
- true_repeated_cross_sections: Logical. Indicates whether the data is truly repeated cross-sections.
- n: The number of observations (units).
- nG: The number of groups.
- nT: The number of time periods.
- tlist: A vector containing all time periods.
- glist: A vector containing all groups.
- call: The call that generated the DIDparams object.

See Also

pre_process_cdid

fonction_simu_attrition

Simulate Unbalanced Panel Data

Description

This function generates a simulated dataset with treatment assignment, individual-level heterogeneity, and time-varying effects. It incorporates attrition based on individual characteristics and time periods. For more details on the methodology, see: Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

```
fonction_simu_attrition(
   N,
   TT,
   theta2_alpha_Gg,
   lambda1_alpha_St,
   sigma_alpha,
   sigma_epsilon,
   tprob
)
```

Arguments

N	Number of units
тт	Number of periods
theta2_alpha_G	g
	Coefficient for interaction between individual heterogeneity and time in the propensity score.
lambda1_alpha_	St
	Coefficient for individual heterogeneity in the propensity score.
sigma_alpha	Standard deviation of individual heterogeneity (alpha).
sigma_epsilon	Standard deviation of the error term (epsilon).
tprob	Probability target to get approximately NTT tprob observations

Value

A data frame containing simulated data.

Examples

```
data_sim <- fonction_simu_attrition(N=150,TT=9,theta2_alpha_Gg = 0.01,
lambda1_alpha_St = 0.5, sigma_alpha = 2, sigma_epsilon = 0.5, tprob=0.5)
```

gg	88	

Description

Function to simplify weight computations. For more details on the methodology, see: Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

Usage

gg(x, thet)

Arguments

x	predictors
thet	parameters

Value

A numeric vector representing the computed weights based on the predictors and parameters.

Examples

```
predictors <- matrix(c(1, 2, 3, 4), ncol = 2)
parameters <- matrix(c(0.5, -0.5), ncol = 1)
gg(predictors, parameters)</pre>
```

gmm_compute_delta_att GMM_compute_delta_att

Description

Function to compute the delta ATT. For more details on the methodology, see: Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

Usage

gmm_compute_delta_att(dp)

Arguments dp

a dp object

Value

a DIDparams object

gmm_convert_delta_to_att

GMM_convert_delta_to_att

Description

Function to process arguments passed to the main methods in the cdid package to compute ATT from deltaATT. For more details on the methodology, see: Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

Usage

gmm_convert_delta_to_att(dp)

Arguments

dp a dp object

Value

a DIDparams object

10

gmm_convert_result GMM_convert_result

Description

Function to convert results so they can be used by the did package developed by Brantly Callaway. For more details on the methodology, see: Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

Usage

```
gmm_convert_result(dp, type)
```

Arguments

dp	a dp object
type	1 for 2step weighting, 2 for identity weighting

Value

a **DIDparams** object

Description

Multi-period objects that hold results for group-time average treatment effects

```
MP(
  group,
  t,
  att,
  V_analytical,
  se,
  с,
  inffunc,
  n = NULL,
  W = NULL,
 Wpval = NULL,
  aggte = NULL,
  alp = 0.05,
 DIDparams = NULL,
  debT
)
```

Arguments

group	which group (defined by period first treated) an group-time average treatment effect is for
t	which time period a group-time average treatment effect is for
att	the group-average treatment effect for group group and time period t
V_analytical	Analytical estimator for the asymptotic variance-covariance matrix for group- time average treatment effects
se	standard errors for group-time average treatment effects. If bootstrap is set to TRUE, this provides bootstrap-based se.
с	simultaneous critical value if one is obtaining simultaneous confidence bands. Otherwise it reports the critical value based on pointwise normal approximation.
inffunc	the influence function for estimating group-time average treatment effects
n	the number of unique cross-sectional units (unique values of idname)
W	the Wald statistic for pre-testing the common trends assumption
Wpval	the p-value of the Wald statistic for pre-testing the common trends assumption
aggte	an aggregate treatment effects object
alp	the significance level, default is 0.05
DIDparams	a DIDparams object.
debT	first time period

Value

MP object

pre_process_cdid Process cdid Function Arguments

Description

Function to process arguments passed to the main methods in the cdid package as well as conducting some tests to ensure data is in proper format and provides helpful error messages.

```
pre_process_cdid(
    yname,
    tname,
    idname,
    gname,
    xformla = NULL,
    data,
    panel = TRUE,
    allow_unbalanced_panel,
```

pre_process_cdid

```
control_group = c("nevertreated", "notyettreated"),
anticipation = 0,
weightsname = NULL,
alp = 0.05,
bstrap = FALSE,
cband = FALSE,
biters = 1000,
clustervars = NULL,
est_method = "dr",
base_period = "varying",
print_details = FALSE,
pl = FALSE,
cores = 1,
call = NULL
)
```

yname	The name of the outcome variable
tname	The name of the column containing the time periods
idname	The individual (cross-sectional unit) id name
gname	The name of the variable in data that contains the first period when a particular observation is treated. This should be a positive number for all observations in treated groups. It defines which "group" a unit belongs to. It should be 0 for units in the untreated group.
xformla	A formula for the covariates to include in the model. It should be of the form ~ X1 + X2. Default is NULL which is equivalent to xformla=~1. This is used to create a matrix of covariates which is then passed to the 2x2 DID estimator chosen in est_method. X's are assumed fixed across the time dimension in this version. Use different columns Xt, Xt+1 if time-varying covariates are needed.
data	The name of the data.frame that contains the data
panel	(Not used) This is not used as balanced and unbalanced panel data is treated similarly.
allow_unbalance	ed_panel
	(Not used) This is not used as balanced and unbalanced panel data is treated similarly.
control_group	Which units to use the control group. The default is "nevertreated" which sets the control group to be the group of units that never participate in the treatment. This group does not change across groups or time periods. The other option is to set group="notyettreated". In this case, the control group is set to the group of units that have not yet participated in the treatment in that time period. This includes all never treated units, but it includes additional units that eventually participate in the treatment, but have not participated yet.
anticipation	(Not used) The number of time periods before participating in the treatment where units can anticipate participating in the treatment and therefore it can affect their untreated potential outcomes

weightsname	The name of the column containing the sampling weights. If not set, all observations have same weight.
alp	the significance level, default is 0.05
bstrap	Boolean for whether or not to compute standard errors using the multiplier boot- strap. If standard errors are clustered, then one must set bstrap=TRUE. Default is TRUE (in addition, cband is also by default TRUE indicating that uniform confi- dence bands will be returned. If bstrap is FALSE, then analytical standard errors are reported.
cband	Boolean for whether or not to compute a uniform confidence band that covers all of the group-time average treatment effects with fixed probability 1-alp. In order to compute uniform confidence bands, bstrap must also be set to TRUE. The default is TRUE.
biters	The number of bootstrap iterations to use. The default is 1000, and this is only applicable if bstrap=TRUE.
clustervars	A vector of variables names to cluster on. At most, there can be two variables (otherwise will throw an error) and one of these must be the same as idname which allows for clustering at the individual level. By default, we cluster at individual level (when bstrap=TRUE).
est_method	the method to compute group-time average treatment effects. At the moment, one can only use the IPW estimator with either "2-step" or "Identity" weighting matrix to aggregate Delta ATT into ATT. include "ipw" for inverse probability weighting and "reg" for first step regression estimators.
base_period	(Not used) The cdid package only uses the g-1 base period for the moment. Whether to use a "varying" base period or a "universal" base period. Either choice results in the same post-treatment estimates of ATT(g,t)'s. In pre-treatment periods, using a varying base period amounts to computing a pseudo-ATT in each treatment period by comparing the change in outcomes for a particular group relative to its comparison group in the pre-treatment periods (i.e., in pre- treatment periods this setting computes changes from period t-1 to period t, but repeatedly changes the value of t)
	A universal base period fixes the base period to always be (g-anticipation-1). This does not compute pseudo-ATT(g,t)'s in pre-treatment periods, but rather reports average changes in outcomes from period t to (g-anticipation-1) for a particular group relative to its comparison group. This is analogous to what is often reported in event study regressions.
	Using a varying base period results in an estimate of ATT(g,t) being reported in the period immediately before treatment. Using a universal base period nor- malizes the estimate in the period right before treatment (or earlier when the user allows for anticipation) to be equal to 0, but one extra estimate in an earlier period.
print_details	Whether or not to show details/progress of computations. Default is FALSE.
pl	Whether or not to use parallel processing
cores	The number of cores to use for parallel processing
call	(Not used) a call control var

print.MP

Value

a DIDparams object

References

Bellego, Benatia, and Dortet-Bernadet (2024), "The Chained Difference-in-Differences", Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2023.11.002.

print.MP

print.MP

Description

Prints a summary of the results contained in an MP object. This function calls summary. MP to display the details of the multi-period analysis results in a user-friendly format.

Usage

S3 method for class 'MP'
print(x, ...)

Arguments

х	An MP object, representing the results of multi-period analysis.
	Additional arguments passed to summary.MP.

Value

No return value. This function is called for its side effects of printing the summary of the MP object to the console.

See Also

summary.MP

Examples

```
# Assuming `mp_object` is a valid MP object
# print.MP(mp_object)
```

process_attgt_gmm Process Results

Description

Process Results

Usage

process_attgt_gmm(attgt.list)

Arguments

attgt.list list of results

Value

list with elements:

group	which group a set of results belongs to
tt	which time period a set of results belongs to
att	the group time average treatment effect

summary.MP

summary.MP

Description

Prints a detailed summary of an MP object. The function outputs key details of the group-time average treatment effects, such as estimation method, control group, and pre-test results for parallel trends.

Usage

S3 method for class 'MP'
summary(object, ...)

object	An MP object, representing the results of a multi-period analysis.
	Additional arguments passed to the function.

summary.MP

Value

No return value. This function is called for its side effects of printing a summary of the MP object to the console, including:

- Call: The call used to create the MP object.
- Group-Time Average Treatment Effects: A table of estimates with confidence bands.
- Control Group: Information about the chosen control group (e.g., "Never Treated").
- Anticipation Periods: Number of periods used to account for anticipation effects.
- Estimation Method: Method used for treatment effect estimation.
- Pre-Test Results: p-values for the test of parallel trends assumption, if available.

See Also

MP, print.MP

Examples

- # Assuming `mp_object` is a valid MP object
- # summary.MP(mp_object)

Index

 $\texttt{att_gt_cdid, 2}$

DIDparams, 5, 10-12, 15

 $\texttt{fonction_simu_attrition, 8}$

gg,9
gmm_compute_delta_att,10
gmm_convert_delta_to_att,10
gmm_convert_result,11

MP, 4, 11, 17

pre_process_cdid, 8, 12
print.MP, 15, 17
process_attgt_gmm, 16

summary.MP, *15*, 16