Package 'triplediff'

July 23, 2025

Title Triple-Difference Estimators

Version 0.1.0

Description Implements triple-difference (DDD) estimators for both average
treatment effects and event-study parameters. Methods include regression
adjustment, inverse-probability weighting, and doubly-robust estimators,
all of which rely on a conditional DDD parallel-trends assumption and
allow covariate adjustment across multiple pre- and post-treatment
periods. The methodology is detailed in Ortiz-Villavicencio and
Sant'Anna (2025) <doi:10.48550/arXiv.2505.09942>.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.3)

RoxygenNote 7.3.2

Imports BMisc (>= 1.4.6), data.table (>= 1.15.0), Matrix (>= 1.6.1),
parallel (>= 1.4.0), parglm (>= 0.1.7), Rcpp (>= 1.0.12),
speedglm (>= 0.3-5)

LinkingTo Rcpp (>= 1.0.12)

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

URL http://marcelortiz.com/triplediff/

NeedsCompilation yes

Author Marcelo Ortiz-Villavicencio [aut, cre],
Pedro H. C. Sant'Anna [aut]

Maintainer Marcelo Ortiz-Villavicencio <marcelo.ortiz@emory.edu>

Repository CRAN

Date/Publication 2025-07-23 19:10:02 UTC

Contents

---

agg_ddd   *Aggregate Group-Time Average Treatment Effects in Staggered Triple-Differences Designs.*

Description

agg_ddd is a function that take group-time average treatment effects and aggregate them into a smaller number of summary parameters in staggered triple differences designs. There are several possible aggregations including "simple", "eventstudy", "group", and "calendar". Default is "eventstudy".

Usage

```
agg_ddd(
  ddd_obj,
  type = "eventstudy",
  balance_e = NULL,
  min_e = -Inf,
  max_e = Inf,
  na.rm = FALSE,
  boot = NULL,
  nboot = NULL,
  cband = NULL,
  alpha = 0.05
)
```

Arguments

ddd_obj   a ddd object (i.e., the results of the [ddd()](#) function)

type   Which type of aggregated treatment effect parameter to compute. "simple" just computes a weighted average of all group-time average treatment effects with weights proportional to group size. "eventstudy" computes average effects across different lengths of exposure to the treatment (event times). Here the overall effect averages the effect of the treatment across the positive lengths of exposure. This is the default option; "group" computes average treatment effects across different groups/cohorts; here the overall effect averages the effect across different groups using group size as weights; "calendar" computes average treatment effects across different time periods, with weights proportional to the group size; here the overall effect averages the effect across each time period.

balance_e	If set (and if one computes event study), it balances the sample with respect to event time. For example, if `balance_e=2`, agg_ddd will drop groups that are not exposed to treatment for at least three periods, the initial period e=0 as well as the next two periods, e=1 and e=2. This ensures that the composition of groups does not change when event time changes.
min_e	For event studies, this is the smallest event time to compute dynamic effects for. By default, `min_e = -Inf` so that effects at all lengths of exposure are computed.
max_e	For event studies, this is the largest event time to compute dynamic effects for. By default, `max_e = Inf` so that effects at all lengths of exposure are computed.
na.rm	Logical value if we are to remove missing Values from analyses. Defaults is FALSE.
boot	Boolean for whether or not to compute standard errors using the multiplier bootstrap. If standard errors are clustered, then one must set boot=TRUE. Default is value set in the ddd object. If boot = FALSE, then analytical standard errors are reported.
nboot	The number of bootstrap iterations to use. The default is the value set in the ddd object, and this is only applicable if boot=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 0.95. In order to compute uniform confidence bands, boot must also be set to TRUE. The default is the value set in the ddd object
alpha	The level of confidence for the confidence intervals. The default is 0.05. Otherwise, it will use the value set in the ddd object.

Value

A object (list) of class agg_ddd that holds the results from the aggregation step.

Examples

```
#----------------------------------------------------------
# Triple Diff with multiple time periods
#----------------------------------------------------------

data <- gen_dgp_mult_periods(size = 500, dgp_type = 1)[["data"]]

out <- ddd(yname = "y", tname = "time", idname = "id",
           gname = "state", pname = "partition", xformla = ~cov1 + cov2 + cov3 + cov4,
           data = data, control_group = "nevertreated", base_period = "varying",
           est_method = "dr")
# Simple aggregation
agg_ddd(out, type = "simple", alpha = 0.10)

# Event study aggregation
agg_ddd(out, type = "eventstudy", alpha = 0.10)

# Group aggregation
agg_ddd(out, type = "group", alpha = 0.10)
```

```
# Calendar aggregation
agg_ddd(out, type = "calendar", alpha = 0.10)
```

---

ddd   *Doubly Robust DDD estimators for the group-time average treatment*
                             effects.

Description

ddd is the main function for computing the Doubly Robust DDD estimators for the ATT, with balanced panel data. It can be used with covariates and/or under multiple time periods. At its core, triplediff employs the doubly robust estimator for the ATT, which is a combination of the propensity score weighting and the outcome regression. Furthermore, this package supports the application of machine learning methods for the estimation of the nuisance parameters.

Usage

```
ddd(
  yname,
  tname,
  idname,
  gname,
  pname,
  xformla,
  data,
  control_group = NULL,
  base_period = NULL,
  est_method = "dr",
  weightsname = NULL,
  boot = FALSE,
  nboot = NULL,
  cluster = NULL,
  cband = FALSE,
  alpha = 0.05,
  use_parallel = FALSE,
  cores = 1,
  inffunc = FALSE,
  skip_data_checks = FALSE
)
```

Arguments

yname	The name of the outcome variable.
tname	The name of the column containing the time periods.

idname	The name of the column containing the unit id.
gname	The name of the column containing the first period when a particular observation is treated. It is a positive number for treated units and defines which group the unit belongs to. It takes value 0 or Inf for untreated units.
pname	The name of the column containing the partition variable (e.g., the subgroup identifier). This is an indicator variable that is 1 for the units eligible for treatment and 0 otherwise.
xformla	The formula for the covariates to be included in the model. It should be of the form ~ x1 + x2. Default is xformla = ~1 (no covariates).
data	A data frame or data table containing the data.
control_group	Valid for multiple periods only. The control group to be used in the estimation. Default is control_group = "notyettreated" which sets as control group the units that have not yet participated in the treatment. The alternative is control_group = "nevertreated" which sets as control group the units that never participate in the treatment and does not change across groups or time periods.
base_period	Valid for multiple periods. Choose between a "varying" or "universal" base period. Both yield the same post-treatment ATT(g,t) estimates. Varying base period: Computes pseudo-ATT in pre-treatment periods by comparing outcome changes for a group to its comparison group from t-1 to t, repeatedly changing t. Universal base period: Fixes the base period to (g-1), reporting average changes from t to (g-1) for a group relative to its comparison group, similar to event study regressions. Varying base period reports ATT(g,t) right before treatment. Universal base period normalizes the estimate before treatment to be 0, adding one extra estimate in an earlier period.
est_method	The estimation method to be used. Default is "dr" (doubly robust). It computes propensity score using logistic regression and outcome regression using OLS. The alternative are c("reg", "ipw").
weightsname	The name of the column containing the weights. Default is NULL. As part of data processing, weights are enforced to be normalized and have mean 1 across all observations.
boot	Logical. If TRUE, the function computes standard errors using the multiplier bootstrap. Default is FALSE.
nboot	The number of bootstrap samples to be used. Default is NULL. If boot = TRUE, the default is nboot = 999.
cluster	The name of the variable to be used for clustering. The maximum number of cluster variables is 1. Default is NULL. If boot = TRUE, the function computes the bootstrap standard errors clustering at the unit level setting as cluster variable the one in idname.
cband	Logical. If TRUE, the function computes a uniform confidence band that covers all of the average treatment effects with fixed probability 1-alpha. In order to compute uniform confidence bands, boot must also be set to TRUE. The default is FALSE.
alpha	The level of significance for the confidence intervals. Default is 0.05.
use_parallel	Logical. If TRUE, the function runs in parallel processing. Valid only when boot = TRUE. Default is FALSE.

cores	The number of cores to be used in the parallel processing. Default is `cores = 1`.
inffunc	Logical. If `TRUE`, the function returns the influence function. Default is `FALSE`.
skip_data_checks	
	Logical. If `TRUE`, the function skips the data checks and go straight to estimation. Default is `FALSE`.

Value

A ddd object with the following basic elements:

ATT	The average treatment effect on the treated.
se	The standard error of the ATT.
uci	The upper confidence interval of the ATT.
lci	The lower confidence interval of the ATT.
inf_func	The estimate of the influence function.

Examples

```
#-----------------------------------------------------------
# Triple Diff with covariates and 2 time periods
#-----------------------------------------------------------
set.seed(1234) # Set seed for reproducibility
# Simulate data for a two-periods DDD setup
df <- gen_dgp_2periods(size = 5000, dgp_type = 1)$data

head(df)

att_22 <- ddd(yname = "y", tname = "time", idname = "id", gname = "state",
              pname = "partition", xformla = ~cov1 + cov2 + cov3 + cov4,
            data = df, control_group = "nevertreated", est_method = "dr")

summary(att_22)


#-----------------------------------------------------------
# Triple Diff with multiple time periods
#-----------------------------------------------------------
data <- gen_dgp_mult_periods(size = 1000, dgp_type = 1)[["data"]]

ddd(yname = "y", tname = "time", idname = "id",
    gname = "state", pname = "partition", xformla = ~cov1 + cov2 + cov3 + cov4,
    data = data, control_group = "nevertreated", base_period = "varying",
    est_method = "dr")
```

---

gen_dgp_2periods	*Function that generates panel data with single treatment date assignment and two time periods.*

Description

Generate panel data with a single treatment date and two periods

Usage

```
gen_dgp_2periods(size, dgp_type)
```

Arguments

size
: Integer. Number of units.

dgp_type
: Integer in {1,2,3,4}. 1 = both nuisance functions correct; 2 = only the outcome model correct; 3 = only the propensity score correct; 4 = both nuisance functions incorrect.

Value

A list with the following elements:

data A data.table in long format with columns:

- id: unit identifier
- state: state variable
- time: time variable
- partition: partition assignment
- x1, x2, x3, x4: covariates
- y: outcome variable
- cluster: cluster ID (no within-cluster correlation)

att True average treatment effect on the treated (ATT), set to 0.

att.unf Oracle ATT computed under the unfeasible specification.

eff Theoretical efficiency bound for the estimator.

gen_dgp_mult_periods   *Generate panel data with staggered treatment adoption (three periods)*

Description

Generate panel data where units adopt treatment at different times across three periods.

Usage

```
gen_dgp_mult_periods(size, dgp_type = 1)
```

Arguments

size   Integer. Number of units to simulate.

dgp_type   Integer in {1,2,3,4}. 1 = both nuisance functions correct; 2 = only the outcome
                model correct; 3 = only the propensity-score model correct; 4 = both nuisance
                functions misspecified.

Value

A named list with components:

data A data.table in long format with columns:

- id: unit identifier
- cohort: first period when treatment is assigned
- partition: partition indicator
- x1, x2, x3, x4: covariates
- cluster: cluster identifier (no within-cluster correlation)
- time: time period index
- y: observed outcome

data_wide A data.table in wide format (one row per id) with columns:

- id, cohort, partition, x1, x2, x3, x4, cluster
- y_t0, y_t1, y_t2: outcomes in periods 0, 1, and 2

ES_0_unf Unfeasible (oracle) event-study parameter at time 0.

prob_g2_p1 Proportion of units with cohort == 2 and eligibility in period 1.

prob_g3_p1 Proportion of units with cohort == 3 and eligibility in period 1.

Index