Package: cdid (via r-universe)

November 8, 2026

Type Package

Title The Chained Difference-in-Differences

Version 1.2.0.1

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

Description Estimates group-time average treatment effects in panels with missing outcomes by the chained difference-in-differences of Bellégo, Benatia and Dortet-Bernadet (2025) <doi:10.1016/j.jeconom.2024.105783>. It is a layer on the 'did' package that keeps the att_gt() and aggte() interface and adds an estimation mode. Chained estimation, the default for panel data, combines short within-unit comparisons by the generalized method of moments. It provides chained inverse probability weighting (IPW), outcome regression (OR) and doubly robust (DR) estimation, where OR and DR use separate or pooled regressions of the controls' outcome changes; custom IPW callbacks; correction for covariate-dependent missing outcomes by observation weighting, change imputation or their augmented combination; parallel trends from each cohort's reference period (PT-post) or from a chosen common start date (PT-all); and a chained conditional pretrend test through conditional_did_pretest(). Ordinary mode delegates unchanged to the Callaway and Sant'Anna (2021) <doi:10.1016/j.jeconom.2020.12.001> estimators in 'did'.

License GPL-3

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

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

Depends R (>= 4.1.0)

Imports did (>= 2.5.1), DRDID (>= 1.4.0), parallel

Suggests testthat (>= 3.0.0), digest, ggplot2, knitr, rmarkdown, generics, jsonlite, waldo, withr

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

NeedsCompilation yes

Author David Benatia [cre, aut], Christophe Bellégo [aut], Joel Cuerrier [aut], Vincent Dortet-Bernadet [aut]

Config/pak/sysreqs libicu-dev

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

Date/Publication 2026-11-07 18:19:10 UTC

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

RemoteRef HEAD

RemoteSha bdad198f120d2d7559a1f6b5738edbad376bb324

Contents

cdid-package 2
aggte 3
att_gt 5
conditional_did_pretest 20
reexports 23

Index **25**

`cdid-package` *cdid: The Chained Difference-in-Differences*

Description

cdid 1.2.0.0, an update of the package after version 0.1.1, is a small layer on the did package. It keeps did's `att_gt()` and `aggte()` interface and adds an estimation mode for panel data with missing outcomes.

- `mode = "ordinary"` delegates unchanged to `did::att_gt()` and `did::aggte()`. This is also the default for repeated cross sections.
- Chained estimation (Bellego, Benatia and Dortet-Bernadet, 2025) is the default for panel data. Built-in doubly robust `est_method="dr"` (the default), `est_method="ipw"`, and control-mean outcome regression `est_method="reg"` are supported, with baseline (`selection = "none"`) or conditional-MTAR (`"conditional_mtar"` / legacy `"corollary1"`) missing-outcome handling via observation weighting (`selection_est_method = "ipw"`), change imputation (`"imputation"`) or the augmented correction (`"dr"`). Custom panel IPW callbacks support baseline selection and, when they return `log_weight_gradient`, conditional-MTAR observation weighting (`selection_est_method = "ipw"`).

cdid requires did 2.5.1 or later and DRDID 1.4.0 or later, and has been tested with did 2.5.1 and DRDID 1.4.0. Later versions are accepted without a guarantee that everything works with them or that results are unchanged. The optional pinned setup in '`validation/SETUP.md`' of the source repository installs exactly the tested versions and verifies their library paths. Use `citation("cdid")` for references.

Details

Quick navigation for offline help:

- att_gt: estimate group-time average treatment effects from long panel data (data) and an optional study-population unit table (frame). Returns a cdid_MP object in chained mode.
- aggte: aggregate a fitted result into "simple", "group", "dynamic" or "calendar" summaries without refitting (cdid_AGGTEobj in chained mode).
- conditional_did_pretest: run the chained conditional pretrend test on panel data (or **did**'s conditional_did_pretest() in ordinary mode).
- reexports: tidy, glance and ggdid generics re-exported from **did**, with chained S3 methods that report cell and bin availability and support reasons.

Run vignette("cdid", package = "cdid") for a self-contained introductory walkthrough. Extended documentation on panel data preparation, investigating missingness and observation models is available at https://github.com/dvbn/cdid.

See Also

att_gt, aggte, conditional_did_pretest, reexports, vignette("cdid", package = "cdid")

---

aggte *Aggregate group-time average treatment effects*

Description

aggte() has the signature of did::aggte() (did 2.5.1). Results from ordinary mode are aggregated by did::aggte() unchanged, with only the arguments you supply forwarded. Chained results (cdid_MP objects from att_gt with est_method = "dr", est_method = "ipw", est_method = "reg", or a custom IPW callback) are aggregated from their retained state without refitting the estimator, with did's selector and inheritance conventions.

Usage

```
aggte(
  MP,
  type = "group",
  balance_e = NULL,
  min_e = -Inf,
  max_e = Inf,
  na.rm = FALSE,
  bstrap = NULL,
  biters = NULL,
  cband = NULL,
  alp = NULL,
  clustervars = NULL
)
```

Arguments

`MP`, `type`, `balance_e`, `min_e`, `max_e`, `na.rm`, `bstrap`, `biters`, `cband`, `alp`, `clustervars`

As in `did::aggte()`: an `MP` object from `att_gt`, the aggregation type (`"simple"`, `"group"`, `"dynamic"` or `"calendar"`), event-time restrictions, missing-value handling and inference settings. For chained results, `NULL` inherits the estimation request's `bstrap`, `biters`, `cband`, `alp` and `clustervars`; explicit values override them with their released did 2.5.1 meaning (including `bstrap = FALSE`, `cband = TRUE`: analytic standard errors with a multiplier band for group/dynamic/calendar bins and did's warning; simple aggregation has no bin band). Cluster overrides that cannot be honoured are refused instead of silently changed. `na.rm = TRUE` requests the explicitly labelled restricted target over the supported cells; with `na.rm = FALSE` an aggregate needing an unavailable cell is `NA` with the missing cells named. As in did, a result fitted with `compute_inffunc = FALSE` has no influence functions and is refused (`"cdid_chained_request_no_inffunc_error"`); multiplier draws of a chained result reuse the workers admitted by its `pl`/`cores` request, with identical results.

Value

For ordinary results, the `AGGTEobj` returned by `did::aggte()`; its `call` element records the delegated call. For chained results, a `cdid_AGGTEobj` (inheriting `AGGTEobj`) with the population cohort-share aggregation (first-calendar cohort masses for reference-date or changing-weight fits on every chained route; see `att_gt`), complete full-frame influence functions including the share uncertainty, the requested interval law, the record that complete per-fit error certification was not requested or performed, and the intended/used/excluded cell records. Multiplier draws use the caller's RNG stream (`set.seed()` before the call reproduces them); no fixed seed exists. Use `summary()`, `tidy`() (columns `label`, `available`, `restricted`, `n.intended`, `n.used`, `excluded.ids`, `reason`), `glance`(), `overall.att`/`overall.se` and `ggdid`() (for `"group"`, `"dynamic"` and `"calendar"` types; `"simple"` has no plot) to inspect aggregated results.

See Also

`att_gt`, `conditional_did_pretest`, `tidy`, `glance`, `ggdid`, `vignette("cdid", package = "cdid")`

Examples

```
set.seed(1)
out <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                    gname = "first.treat", data = did::mpdta,
                    mode = "ordinary", bstrap = FALSE)
cdid::aggte(out, type = "simple")
cdid::aggte(out, type = "group")
```

att_gt *Group-time average treatment effects: chained or ordinary mode*

Description

`att_gt()` has the interface of `did::att_gt()` (did 2.5.1) plus a `mode` argument, the arguments that configure chained estimation and one chained-only advanced `control` list.

`mode = "ordinary"` passes the call unchanged to `did::att_gt()` and returns its `MP` object. For panel data, the default `mode = NULL` selects chained estimation. Chained estimation runs for built-in doubly robust `est_method = "dr"` (the default), explicit built-in outcome regression `est_method = "reg"`, built-in `est_method = "ipw"` (chained IPW-GMM, Bellego, Benatia and Dortet-Bernadet 2025) or a did-style custom IPW panel callback function `f(y1, y0, D, covariates, i.weights, inffunc = TRUE, ...)` returning a list with finite scalar `ATT` and finite length-m_ℓ vector (or 1-column matrix) `att.inf.func` in the supplied row order. Unsupported `est_method` strings are refused with an explanatory error before the data is touched; built-in IPW is never substituted for them. For repeated cross sections (`panel = FALSE`) the default resolves to ordinary mode.

Usage

```
att_gt(
  yname,
  tname,
  idname = NULL,
  gname,
  xformla = NULL,
  data,
  panel = TRUE,
  allow_unbalanced_panel = NULL,
  control_group = c("nevertreated", "notyettreated"),
  anticipation = 0,
  weightsname = NULL,
  fix_weights = NULL,
  alp = 0.05,
  bstrap = TRUE,
  cband = TRUE,
  biters = 1000,
  clustervars = NULL,
  est_method = "dr",
  base_period = "varying",
  faster_mode = TRUE,
  print_details = FALSE,
  pl = FALSE,
  cores = 1,
  compute_inffunc = TRUE,
  ...,
  mode = NULL,
```

```
  frame = NULL,
  selection = "none",
  selection_model = "pair_logit",
  selection_pooling = "separate",
  selection_xformla = NULL,
  selection_universe = "F",
  selection_rho = NULL,
  selection_justification = NULL,
  selection_est_method = "ipw",
  imputation_pooling = "separate",
  gmm_weighting = c("optimal", "identity"),
  gmm_covariance = "tikhonov",
  pt_assumption = c("post", "all"),
  pt_start = NULL,
  control = NULL,
  outcome_pooling = "separate"
)
```

Arguments

yname, tname, idname, gname, xformla, data, panel, control_group, anticipation, weightsname, fix_weights, alp, bstrap, cband, biters, clustervars, est_method, base_period, faster_mode, print_details, pl, cores, compute_inffunc

: As in `did::att_gt()`, with the same names, order and defaults. In ordinary mode, arguments are forwarded only if you supplied them, so did applies its own defaults. Each expression is evaluated once, in your environment, when did needs it. The exception is `panel`: with `mode = NULL`, cdid evaluates it first to choose the mode. The default `est_method = "dr"` is kept. In chained mode `idname` is required (units are aligned by identity) and at most one cluster variable beyond `idname` is accepted. `compute_inffunc = FALSE` returns point estimates only, as in did: `bstrap` and `cband` become `FALSE`, `inffunc`, `V_analytical`, `W` and `Wpval` are `NULL`, `se` is `NA`, `c` is the pointwise normal constant and `aggte()` refuses the result; the feasible two-step/Tikhonov GMM weights still use the complete internal link influence functions, so every estimate equals the `compute_inffunc = TRUE` estimate. `pl = TRUE` with `cores = k` runs the multiplier bootstrap's draw products on `k` worker processes (forked on Unix-alikes, a PSOCK cluster on Windows; base `parallel`), which must be declared through `control = list(workers = k)`; otherwise the request is refused before the data are read. Signs are drawn by your session in the serial batch order, so draws, standard errors and bands equal `cores = 1`, later `aggte()` calls inherit the workers, and estimation itself stays serial. `print_details = TRUE` reports stage and comparison progress through messages (`suppressMessages()` silences it); `FALSE` keeps the fit quiet. `faster_mode` keeps its documented meaning: chained mode has one execution engine. The public never-treated code is `gname == 0`. Chained `fix_weights` accepts `NULL`, `"first_period"` and `"base_period"` (`"varying"` is refused); see 'Reference dates and changing weights' for covariates and sampling weights that change within a unit.

allow_unbalanced_panel
: NULL (default), TRUE or FALSE. In ordinary mode, NULL means did's default FALSE (balance the panel). Chained estimation retains every eligible unit (at least two observed outcomes), including balanced panels, so an explicit FALSE in chained mode is an error that points to mode = "ordinary".

...
: Passed to a custom est_method function in ordinary mode (as in did) and in chained mode when est_method is a function. When est_method is a character string (such as "ipw") in chained mode, non-empty ... is refused.

mode
: NULL (default), "chained" or "ordinary". NULL selects "chained" when panel = TRUE and "ordinary" for repeated cross sections.

selection_universe
: Selection logit fitting sample: "F" (default, all supplied representative units), "B" (at least one observed outcome), or "A" (at least two). "B" and "A" are supported only for selection_model = "pair_logit"; selection_model = "independent_period_logit" requires selection_universe = "F". These outcome counts do not specify structural observation windows. The entire F remains the population and inference frame.

selection_rho
: NULL for rho=1, or a column name in the unit frame containing fixed positive KNOWN normalized rho = f_P(X|G,U)/f_P(X|G) on U. Only supported for A/B under selection_model = "pair_logit"; no estimation or self-normalization. Missing X/rho outside U is allowed and never evaluated. Necessary covariates must be complete on U.

selection_justification
: A nonempty explanation required for A/B. With NULL rho this explicitly asserts U is representative within cohort under the declared weighted population law (U independent of X conditional on G). Otherwise it declares provenance of the known normalized transport. The package records this assertion; fit diagnostics cannot establish it.

selection_model
: Observation-selection probability model when selection correction is active: "pair_logit" (default, direct logit for observing both endpoints of an ordered pair) or "independent_period_logit" (period-observation logit with pair probability formed as the product of the two endpoint date probabilities on selection_universe = "F"). Ignored when selection = "none" unless a non-default value is supplied without an active selection assumption, which is refused. Inactive and never evaluated under selection_est_method = "imputation"; active under selection_est_method = "dr", where it defines the observation-probability arm of the augmented correction.

selection_pooling
: Slope-pooling mode across active selection cells for both "pair_logit" and "independent_period_logit": "separate" (default, cell-specific slopes), "within_cohort" (one shared slope vector per cohort across its active pairs or dates, with cell-specific intercepts when the selection formula includes an intercept), or "across_cohorts" (one shared slope vector across all active cells, with cell-specific intercepts when the selection formula includes an intercept). Pooling restricts selection slopes only; outcome_pooling independently controls counterfactual outcome-regression slopes. Inactive and never evaluated under selection_est_method

= "imputation", which uses imputation_pooling instead. Under selection_est_method = "dr" it pools the observation-probability slopes and is independent of imputation_pooling.

selection_est_method

: Chained-only observation-correction method, separate from the treatment comparison method est_method: "ipw" (default, the existing probability-weighted correction, unchanged), "imputation" (pure actual-change imputation; see the section below) or "dr" (augmented observation correction combining both; see the section below). "imputation" requires selection = "conditional_mtar" (or the legacy assumption alias "corollary1"), selection_universe = "F" with selection_rho = NULL, known cohorts and complete xformla covariates on every F row, and a built-in est_method; it fits no observation-probability model, so selection_model, selection_pooling and selection_xformla are inactive and never evaluated. "dr" has the same declaration requirements, supports built-in est_method = "ipw", "reg" or "dr"; custom callbacks remain unavailable. It uses every probability-model and imputation argument. Treatment IPW fits no counterfactual regression; treatment OR fits no treatment propensity; treatment DR fits both. Ordinary mode neither evaluates nor forwards it.

outcome_pooling

: Chained counterfactual outcome regression for est_method = "reg" or "dr": "separate" (default) fits coefficients for each distinct date-pair/control-group comparison; "pooled" shares slopes per stored-calendar step across comparisons, with local intercepts when xformla includes an intercept. Reference covariates and sampling weights are part of the comparison's fit identity. Identical fits are reused. This setting is independent of selection_pooling and imputation_pooling; it changes neither treatment IPW nor ordinary mode. See 'Control-mean outcome regression and doubly robust estimation'.

imputation_pooling

: Slope pooling of the actual-change imputation model, used only when selection_est_method = "imputation" or "dr" (otherwise it is never evaluated): "separate" (default; one slope vector per actual cohort/pair cell), "within_cohort" (one slope vector per cohort across its needed pairs) or "across_cohorts" (one slope vector across all needed cells). Local cohort/pair intercepts are kept whenever xformla has an intercept. Pooling is an explicit model restriction, never an automatic fallback; it can be wrong, for example when treated cohorts' dynamics differ across pairs. Ordinary mode neither evaluates nor forwards it.

frame, selection, selection_xformla, gmm_weighting, gmm_covariance, pt_assumption, pt_start

: Chained-only: an optional unit frame with population or sampling information (one row per population unit, matched to data by idname; frame-only units keep their cohort, weight and cluster for population shares and influence-function rows), the observation-selection assumption and its covariate formula (selection = "none" by default, or "conditional_mtar" for conditional mean trend-MAR selection correction, with "corollary1" accepted as the legacy direct-pair alias for selection = "conditional_mtar" with selection_model = "pair_logit"; selection_xformla is inactive and never evaluated under selection_est_method = "imputation"; under "dr" it defines only the observation-probability design, NULL inheriting xformla), the GMM weighting ("optimal"

by default for feasible two-step weighting, with "two_step" accepted as an explicit alias for "optimal", and "identity" for explicit unweighted estimation), the singular or near-singular moment-covariance handler under feasible two-step weighting (gmm_covariance = "tikhonov" by default for covariance-ridge Tikhonov regularization; Moore-Penrose is deferred and is rejected), and the separately selected parallel-trends assumption ("post" by default; "all" adds untreated-trend restrictions from an explicit common pt_start). Under "all", pt_start must be one finite numeric date exactly in the stored calendar, chosen exogenously; omission, NULL and non-calendar dates are errors. Under chained "post", a non-NULL start is inapplicable and refused. Ordinary mode delegates before chained-only validation: these arguments are neither evaluated nor forwarded to did. pt_assumption is separate from base_period, whose did meaning and default ("varying") are unchanged.

control Chained-only advanced list, NULL by default. NULL (and an empty list) resolves to the adopted defaults recorded in the result: numeric_workspace_bytes = 1073741824 (1 GiB), workers = 1, and no user estimated-work budget (counted_work_limit = Inf). The estimated-work cap is opt-in: set counted_work_limit to a finite positive whole number (exact binary64 integer) to refuse, before the stage runs, a fit whose conservative preflight work ESTIMATE exceeds it; an explicit Inf also opts out. Zero, negative, NA/NaN, non-integral, out-of-range and non-numeric values are refused. Work estimates are diagnostic and conservative, not measured runtime. Internally a disabled cap hands the finite private constructors the exact-count arithmetic ceiling 2^53 - 1; the result records the public value and that ceiling separately (chained$public$counted_work_limit, chained$public$work_cap). These controls admit counted numeric payload and work estimates; they do not bound process RSS, R heap, CPU time or whole-call work, and cdid does not detect machine RAM or time. A plain named list may also override numeric_workspace_bytes, workers (the number of multiplier bootstrap worker processes pl = TRUE, cores = k may use; the default 1 is never raised automatically and their whole buffers count against the same numeric workspace), chunk_rows, condition_trigger (default 1e-10, finite and strictly between zero and one), and ridge_relative (default 1e-4, finite and positive). Unknown entries and objects are refused. The universal per-fit error guarantee and its relative_accuracy override are deferred; supplying that entry raises cdid_chained_public_accuracy_deferred. Ordinary mode never evaluates it.

Details

PT-all retains every baseline PT-post condition, including at qualified causal dates before the chosen start. Each cohort keeps its original admissible reference date. Added equalities constrain represented operator-specific untreated pre-period differentials from the common start up to that reference; earlier differentials and anticipation effects stay free. The same comparison catalog, control operators, observation pools and complete influence functions are used. Sampling, selection, overlap and nuisance assumptions remain required. If there are no represented extra restrictions globally, the fit is allowed and summary/print reports "No additional restrictions" (the PT-post-equivalent model). Stored window dates and represented restrictions are counted separately by cohort in result$chained$pt; no cohorts are dropped.

Both base_period values retain did's direct, unrestricted pre-period reporting. Restricted zero pre-parameters never supply the full pretrend Wald test. Descriptive pre-start, within-window and anticipation tags accompany the test record. Pre-start contrasts may reject without contradicting PT-all from the chosen start; anticipation effects need not be zero. An in-window auxiliary comparison is implied zero only under its actual control operator's PT and selection premises; a date tag alone does not cover an auxiliary NYT pool absent from the catalog assumptions. There is no new subset or overidentification test.

GMM weighting and covariance handling: feasible two-step GMM weighting (gmm_weighting = "optimal", or alias "two_step") is the default, and gmm_weighting = "identity" is the explicit unweighted option. Proved exact structural equivalence reductions are applied before covariance conditioning; when structural singularity is removed by exact reduction alone and the reduced transformed covariance $B = C\widehat{V}_r C'$ (with $G = E'E = C'C$ from the sparse reduction reconstruction E) is well conditioned ($\rho \geq \tau_{\text{cond}}$), ordinary Cholesky-whitened QR is used without invoking a singular-covariance handler. When B is singular or near-singular ($\rho < \tau_{\text{cond}}$), the selected handler gmm_covariance = "tikhonov" uses covariance ridge with $\lambda = \alpha_{\text{ridge}} \operatorname{tr}(B)/M$ and effective reduced weight $C'(B + \lambda I_L)^{-1}C$, where M is the full catalog size. The trigger uses the reciprocal spectral condition $d_{\min}(B)/d_{\max}(B)$, rather than a 1-norm condition estimate. Known unit-covariance rank deficiency or failed ordinary Cholesky on valid PSD covariance also activates ridge. Identity is never substituted automatically. At exactly zero covariance the criterion is full-catalog least squares (reduced weight G up to scale), with the actual zero sandwich and the existing zero-scale inference diagnostics. No moments, units or identified contrasts are dropped. Internally formed complete-unit-IF Gram roundoff is diagnosed under the existing eigen-residual rule; no eigenvalues are clipped and no penalty is increased after failure. Nonfinite, materially indefinite and invalid provenance inputs still refuse. Tuning, activation, full/retained dimensions, scale, conditioning and provenance are recorded in the result. Analytical and multiplier-bootstrap inference always use the complete target influence functions and the true sandwich variance, never the inverse regularized criterion matrix. Clusters affect inference only; weights and tuning use unit IF covariance. Feasible weighting does not imply finite-sample unbiasedness. Inference requires the complete-link ALR/CLT, covariance consistency and a stable limiting loading/automatic branch; no uniform boundary or zero-population-trace inference guarantee is claimed.

Numerical policy: conventional Cholesky/QR computation uses the existing conditioning and factor-residual policies. Relative solve residuals remain diagnostic, including at zero dual solutions. Complete per-fit interval error certification is not requested or performed; no universal 1e-6-of-SE guarantee is claimed. Its public override is explicitly deferred.

Pretrend test: the full set of declared nonreference pretrend comparisons of the selected base_period reporting is tested with the did-style Wald statistic when every comparison is available and the joint covariance admits the full Cholesky/solve under the existing conditioning policy; otherwise the test is unavailable with its reason and the cells are returned unchanged. It is the unconditional test; the conditional pretrend test is conditional_did_pretest().

The package refuses to load with did older than 2.5.1 or DRDID older than 1.4.0; it has been tested with did 2.5.1 and DRDID 1.4.0, and later versions are accepted but untested. Qualify calls as cdid::att_gt() so that a later library(did) cannot change which function runs.

Input and result inspection: in chained mode, data is a long panel with unique (idname, tname) rows, finite numeric tname, numeric yname (finite when observed and kept as NA when missing, never replaced with 0), and constant unit cohort gname (0 for never-treated units). No unit requires a complete history or an observed outcome at the cohort baseline date. The optional unit table frame defines the study population F of size $N = |F|$, retaining units observed at two or more dates (A),

singletons (one observed outcome) and zero-outcome units in population counts, cohort shares, active selection or imputation models and influence functions for all study units without dropping rows. Under the default `selection = "none"`, units observed at both endpoints of each used date pair are assumed representative of their cohort in the study population with respect to the mean outcome change over that pair. Use `summary()`, `tidy()` (columns `target.id`, `role`, `available`, `support.status`, `support.reason`), `glance()`, `aggte()` and `ggdid()` to inspect fitted results; see `reexports` and `vignette("cdid", package = "cdid")`.

Value

In ordinary mode, the `MP` object returned by `did::att_gt()`, unchanged. did records the delegated call (for example `did::att_gt(yname = yname, ...)`) as `DIDparams$call`, so printed summaries show that call.

In chained mode with `est_method = "dr"`, `est_method = "reg"`, `est_method = "ipw"` or a custom IPW callback function, a `cdid_MP` object (inheriting `MP`, built with did's constructors) whose did fields hold the reported chained cells: direct did-style pretrend comparisons for every declared treated cohort and the ATT(g,t) post cells, with complete full-frame influence functions, the analytic (or multiplier IQR) standard errors and the critical value of the requested interval law. For a custom `est_method` callback, each qualified link invokes `est_method(y1, y0, D, covariates, i.weights, inffunc = TRUE, ...)` on the observed risk set in canonical increasing A-unit order with calibrated weights `a_i = w_i / max(r_G, 1e-6)`, composes cohort-rate estimation uncertainty via `b_k = -sum(psi[G == k]) / (m * r_k)` (zero when `r_k <= 1e-6`) and `phi = (n / m) * 1[observer] * psi + rho %*% b` without adding a duplicate propensity or pre-GMM cohort-share term. A callback may also return the optional `log_weight_gradient`: a finite length-m_ℓ vector (or 1-column matrix) in the supplied row order with $g_j = \partial ATT / \partial \log$ `i.weights[j]`, the total derivative with every internal refit included (ATT units, never centred). When supplied the rate channel uses `b_k = -sum(g[G == k]) / r_k` instead of the identity rule (the two are never added). Under `selection = "conditional_mtar"` (`selection_est_method = "ipw"`, every selection model, pooling and universe) callbacks receive the selection-corrected weights `w a_U / (rho max(p_U, 1e-6))` and must supply this gradient; `b` then loads the selection rates and logit coefficients and `phi = (N / m) * 1[observer] * psi + xi_u %*% b` on every original row. Without it such requests are refused: $g = \psi/m$ needs prerequisites that are not verified and g is never inferred. Shape checks do not establish the callback's scientific validity. Callbacks must reimplement the existing IPW criterion with the same weighted propensity equations, normalized outcome means and complete fixed-rate influence function, including propensity and normalization terms. The IF must have valid mean-zero semantics and preserve the supplied row order; finite return shapes do not establish correctness. The bridge retains the callback IF unchanged, records its mean diagnostically, and imposes no centering cutoff. The existing point/design/complete-IF forest identities apply after outcome-free verification of equal cohort/control pools and observation masks, including balanced panels. Broader arbitrary-callback support requires a separate public policy. Arbitrary workspace inside the opaque callback is outside package-owned array accounting. Cells the retained comparison catalog cannot support are `NA` with a support reason; they are not population non-identification claims. The `chained` element records the support ledger, the GMM weighting and covariance-handling record (`covariance_handling`), the record that complete per-fit error certification was not requested or performed, the pretrend test (`W`/`Wpval` are set only when the full declared test is computable; otherwise an explicit unavailable or not-applicable status and reason), the resolved resource profile with its status, the actual input dimensions and derived finite bounds, the resource accounting of reachable numeric payload (not process memory or CPU), and the state later aggrega-

tion needs so that aggte never refits. The resource accounting records the exact auxiliary reporting plan (auxiliary_plan: the union of the varying and universal pretrend comparisons that are fitted, the reported pretrend columns of the chosen base_period, reference-zero and unavailable cells), the mask registry's charged scans and retained records (mask_registry) and, on the built-in IPW reduction route, whether the registry's sealed mask-equivalence certificate replaced streamed mask comparisons (early_structural$mask_evidence). The inherited flags as honoured, the admitted multiplier workers and their route are recorded in chained$request (execution); DIDparams carries the requested print_details, pl, cores and compute_inffunc.

Refusals are explicit errors inheriting "cdid_error":

- "cdid_chained_unavailable", also inheriting "cdid_chained_request_est_method", for unsupported string chained requests.
- "cdid_chained_request_unsupported_selection" for unsupported missing-outcome correction requests, including custom callbacks with imputation or the augmented correction.
- "cdid_custom_ipw_error" for malformed or nonfinite custom callback outputs.
- "cdid_chained_panel" if panel is not TRUE.
- "cdid_chained_balanced" for an explicit allow_unbalanced_panel = FALSE.
- "cdid_chained_public_control_error" for an invalid control.
- "cdid_chained_request_parallel_error" for pl = TRUE, cores = k beyond the declared control$workers.
- "cdid_chained_request_*" for unsupported requests or inputs.
- The accepted pipeline's typed bound errors when the input's conservative numeric-payload or work envelope exceeds the profile (nothing is dropped or relaxed), and genuine invalid nuisance, nonfinite, covariance or solve errors.

Exact signed-coordinate separation is a sufficient finite-root obstruction; it is not a complete separation classifier or a population identification conclusion. An invalid mode, or an invalid panel when resolving mode = NULL, raises "cdid_mode_error".

Reference dates and changing weights (chained mode)

Calls whose covariates and sampling weights pass the existing static checks (constant within unit and in agreement with frame) run unchanged, and fix_weights = NULL, "first_period" and "base_period" give identical results. Sampling weights that change within a unit require "first_period" (every context uses w at the first stored calendar date t_1) or "base_period" (every unit of target cohort g's context, controls included, uses w at $b_g = \max\{d : d + \texttt{anticipation} < g\}$); NULL would mean did's period-specific cell weights and is refused. With unit-level covariates and w_{t_1} present on every study unit, "first_period" gives the existing static estimator on w_{t_1}, on every route.

Covariates that change within a unit (or a formula use of tname) are reference-date covariates, supported on every chained route: baseline IPW, custom did-style callbacks, OR and DR (selection = "none"); conditional selection with IPW, OR or DR (pair_logit on selection_universe F, B or A with known selection_rho, independent_period_logit on F, every selection_pooling); pure actual-change imputation with IPW, OR or DR (every imputation_pooling); and the augmented observation correction (selection_est_method = "dr"). Every post comparison of target g (bridges and controls included) uses each unit's values at b_g; an auxiliary pre-period report

uses the earlier endpoint of its date pair. Values come only from the actual (id, date) rows, including rows with a missing outcome, and are required only where an estimate reads them: the outcome/treatment design on observed pairs (for every study unit in each cohort for pure imputation and the augmented correction), the selection design in the selection fitting sample, and W_g on each rate or mass denominator. A missing required value is refused with its id, date, variable and support (never carried forward or substituted). Each formula, its factor levels and transform state are evaluated once on the canonical required (date, id) rows of all post and both auxiliary report families; `selection_xformla` keeps its own design. Rates, repeat rates, selection cells, treatment logits, imputation cells and counterfactual regression blocks are keyed by recipe (design date, weight anchor) and fitted once. Separate outcome regressions keep block-specific slopes; `outcome_pooling="pooled"` shares a global slope on $L_e z$. Selection or imputation pooling never crosses recipes. Custom callbacks receive the recipe design and W_g/r_G. Every influence function has one row per original unit with all rate, selection, imputation, treatment and outcome-regression channels. A frame value cannot override data values that change. On these fits the dated baseline IPW and custom routes use the general GMM criterion over every catalog comparison rather than the static structural reduction.

For these reference-input fits, `aggte()` weights cohorts by their first-calendar masses $H(I_g w_{t_1})$, with share uncertainty in the influence functions; never-treated units need no w_{t_1}, and an aggregate whose treated cohort lacks w_{t_1} for some unit is `NA` with the ids. Each cell keeps its context weight; with `"base_period"` an aggregate is a declared mixture of cohort effects, not the ATT of one weighted population.

Conditional MTAR selection correction

Use `selection="conditional_mtar"` (or the legacy direct-pair alias `selection="corollary1"`) with built-in `est_method="dr"`, `est_method="ipw"`, or `est_method="reg"`. `selection_xformla=NULL` inherits `xformla`; both formulas are one-sided, fixed unit-level designs preserving `stats::terms`, `assign` and intercept metadata, including intercept-only (`~1`), no-intercept (`~0 + x`) specifications. Only `selection_xformla` supports a zero-column design (`~0`, fixed cell probability 0.5 and date product 0.25 with zero parameters); `xformla` requires a nonempty design. `selection_model="pair_logit"` (default) fits the joint observation indicator for each active ordered pair and cohort; `selection_model = "independent_period_logit"` fits period-observation logits on `selection_universe="F"` and forms each pair probability as the product of its two endpoint date probabilities without A/B multiplication. Both models support `selection_pooling="separate"` (default), `"within_cohort"` and `"across_cohorts"` over active union-catalog cells. The legacy alias `selection="corollary1"` is restricted to `selection_model="pair_logit"`. Identical actual models are shared across NT/NYT pools and direct pre-comparisons, and reused date or pair scores are summed by original unit before covariance computation. This is logit only.

F is the supplied representative unit table, in frame order. When frame is omitted, the units listed in the supplied panel data, including singletons, define F; no wider population register is presumed. External linked frame-only rows may have zero outcomes but need known cohort, positive fixed weight (if weighted), necessary covariates and cluster labels. An absent outcome is a known pair nonresponse only under the user's linkage/observation coverage. No frame rows are dropped. Singletons supply no differences.

For U=F, the combined correction is a/p. Retained U=A/B uses its own conditional logit p_U and repeat rate a_U with b=a_U/(rho*p_U), the fitted probability (or date product) floored at 1e-6 as described under Probability clipping. IPW uses b in the treatment score and both empirical outcome means; outcome regression uses b in the control-only weighted regression and the target residual

mean. The a_U denominator is U; for A, a_U=1 and its rate influence is zero. Known rho contributes no nuisance influence. Its normalization is a scientific input, not a sample renormalization. F-logit correctness does not imply U-logit correctness; these are different modeling assumptions, without an efficiency ranking.

Every actual catalog pair, including pre-comparisons, needs its own population conditional mean trend-MAR restriction. Adjacent-pair MAR is insufficient. Selection and, for IPW, synthetic treatment-odds specifications must be correct conditional on all scientific X; OR requires the correct control-mean regression; different predictor lists require separately sufficient restrictions. PT/no anticipation, weighted-law validity, positivity, finite regular full-rank interior roots and complete-score second moments also remain application assumptions. No trimming, imputation or new nuisance regularizer is used; the probability clipping is described in its own section.

Complete influence functions replay every F row, including nonobserver selection/rate scores. Public n is N=|F|. The GMM core uses psi=(n_A/N)*phi_F on all F rows with its original n_A covariance scale. Population shares, share/ATT cross covariance and all cluster rows are retained, including clusters with no A units. Clustering changes inference only. Existing GMM asymptotic loadings apply; they are not exact case derivatives of refitted two-step covariance weights. Full requested pretrend nulls retain their computed result or explicit unavailable reason.

Known pair probabilities (including design p=1), custom-IPW callbacks without an explicit `log_weight_gradient`, estimated/unnormalized transport and probit are unsupported. All-success data never establish a design probability: fitted logit refuses separation/boundary/rank failures. Numerical certificates are not provided for this branch. Finite tests and the demo do not establish inference coverage.

Control-mean outcome regression and doubly robust estimation

Explicit `est_method="reg"` uses control-mean outcome-change regression. Default or explicit `est_method="dr"` combines that regression with residualized Hajek IPW using the risk-set treatment logits from `xformla`. Treatment and selection designs retain their own columns; outcome alias preprocessing applies only to the outcome regression. Treatment OR fits no treatment propensity or treated outcome regression.

With `outcome_pooling="separate"` (default), each unique comparison block b has its own slope vector β_b and, when the formula includes an intercept, its own intercept α_b. A block is a date pair and control group with its reference covariates, weights and nuisance context. Repeated requests for the same fit reuse it. Explicit `outcome_pooling="pooled"` fits one shared slope vector β with local block intercepts, preserving the earlier outcome model. Both fits use the qualified catalog and auxiliary comparisons.

The design is $d_{e,i} = L_e z_i$, where $L_e = i_t - i_s$ counts transitions between stored-calendar positions, never raw subtraction of numeric date labels. In separate mode L_e is constant within a fit, so it only rescales that block's coefficients. Pooled mode instead restricts covariate slopes to be proportional to this duration. Pooling is an outcome-model restriction, not an additional parallel-trends assumption.

For baseline and observation-weighted OR, the target link is the normalized weighted target-cohort residual mean $\hat{\mu}_\ell = \sum_{i \in S_\ell} q_{\ell,i}(\Delta Y_{e,i} - \hat{\alpha}_{b(\ell)} - L_{e(\ell)} z_i' \hat{\beta}_{b(\ell)}) / \sum_{i \in S_\ell} q_{\ell,i}$, with a common $\hat{\beta}$ in pooled mode. Baseline weights divide fixed unit weights by each cohort's pair response rate among repeat observers. Conditional observation weighting retains the declared probability model, fitting population and known transport information. Imputation and augmented correction apply the same counterfactual regression choice to reconstructed actual changes; their actual-change models remain controlled independently by `imputation_pooling`.

Actual `stats::terms` and `assign` metadata determine intercepts and retained columns. Intercept-only `~1` uses block means without slopes; `~0+x` uses uncentered slopes without intercepts. An explicitly named `` `(Intercept)` `` variable is a slope column when the formula has no intercept. No intercept is silently added.

Standard pivoted QR (default tolerance `1e-7`) proposes aliased columns. Each proposal must reconstruct every required control and treated prediction row at arithmetic roundoff (`64*.Machine$double.eps` times the original column count). A nontrivial discrepancy retains the original column. The reduced regression still must pass its rank and conditioning checks. Separate fits check each block's prediction support; pooled fits can identify a shared slope across blocks. There is no automatic switch from separate to pooled, regularization of the outcome regression, or removal of observations or moments to obtain a fit. Removing a redundant column changes the parameterization, not an assumption about its original coefficient.

Static outcome covariates are needed on repeat observers for baseline and observation-weighted routes. Dated designs use the required reference rows. Imputation and augmented correction require their designs over the relevant complete unit table. Every original unit retains all its nuisance contributions, including units with zero or one observed outcome. Cluster inference sums these original-unit influence functions. Separate coefficients do not make fits on overlapping units statistically independent.

The fitted mode and compact regression records are reported in `chained$selection` and `chained$nuisance_diagnostics`, with retained bases and duration conventions. OR/DR influence functions include the estimated outcome coefficients and all active observation, imputation, rate and treatment nuisance paths. All original comparisons, GMM, aggregation, multiplier inference and pretrend reporting remain. Support or conditioning failures of an active regression are computation failures; they do not by themselves prove population ATT nonidentification. No numerical accuracy certificate or additional public outcome-regression tolerance is introduced.

Pure actual-change imputation

With `selection = "conditional_mtar"` and `selection_est_method = "imputation"` the observation correction predicts instead of weighting. For every actual cohort/pair cell (k, e) needed by the union of the post catalog and both base-period auxiliary catalogs, the actual-change model $j_{k,e}(X) = \alpha_{k,e} + L_e z(X)'\beta_{r(k,e)}$ is fitted by weighted least squares to the observed pair changes ($S_e = 1$) of cohort k with the fixed unit weights (no observation probabilities), with one local intercept per cell when `xformla` has an intercept and slopes shared as `imputation_pooling` declares. Each fitted $j_{k,e}$ is evaluated on every F row of cohort k (prediction standardization, not a fill-in of the unobserved changes only), with the empirical repeat rates a_k on F.

The treatment method then uses these reconstructions: `est_method = "ipw"` fits the risk-set treatment propensity logits with synthetic weights wa_k and no counterfactual outcome regression; `est_method = "reg"` fits the counterfactual control outcome-change regression to the imputed synthetic control changes, with separate or shared slopes according to `outcome_pooling` and local intercepts when declared by `xformla`, and no treatment logit; `est_method = "dr"` fits both. The counterfactual regression is a second model, distinct from the actual-change imputation model, although both use `xformla`.

`xformla` is evaluated on every F row, including frame-only, singleton and zero-observation units; no common repeat-observer alias reduction is applied. Each imputation pooling group and each counterfactual regression group apply their own QR alias/support preprocessing on their fitting rows with prediction-row alias verification; the treatment logit keeps the raw columns with the

existing rank and separation refusals. Intercept-only and no-intercept formulas are supported; empty formulas are refused. A rank-deficient pooling group refuses; nothing is pooled automatically.

Sufficient (not necessary) conditions: actual-pair conditional mean trend-MAR on F for every needed cell, positive cohort and cell masses, a correct linear actual-change mean within each pooling group, and in addition a correct synthetic treatment logit (`"ipw"`), a correct counterfactual control-mean regression in the relevant block (`"reg"`) or either of the two (`"dr"`), with PT/no anticipation and the fixed-weight population law. When `outcome_pooling="pooled"`, the counterfactual model must be correct across its shared blocks. For actual-change `imputation_pooling`, `"separate"` is the flexible default; `"within_cohort"` and `"across_cohorts"` are explicit restrictions that can be misspecified, for example under heterogeneous treated dynamics. No observation-probability model, date-independence factorization or observed-level conditioning is used, and no refusal arises merely because an unused observation logit would be separated (for example on a balanced panel).

Complete original-unit influence functions on all F rows, including off-A and frame-only units, carry the repeat-rate and imputation paths and the active treatment/counterfactual nuisance paths into the existing GMM, aggregation, clustering, multiplier-band and pretest consumers with core scaling $\psi = (n_A/N)\phi_F$. `chained$selection` reports `method = "imputation"`, both method axes, the pooling and the inactive arguments; `chained$nuisance_diagnostics` reports the cells, pooling groups (ranks, slopes, aliases), counterfactual regression, treatment logits, fitted catalog and study-population support with zero observation-probability models; `chained$outcome_design` records the raw full-F design. B/A fitting samples, transport columns and custom IPW callbacks are unavailable; the augmented correction that adds an observation-probability arm is described in the next section. No numerical certificate or coverage claim is made.

Augmented observation correction

With `selection = "conditional_mtar"` (or the legacy direct-pair alias `"corollary1"`), `selection_est_method = "dr"` and built-in `est_method = "ipw"`, `"reg"` or `"dr"`, every needed actual cohort/pair cell $c = (k, e)$ is reconstructed by

$$\widehat{Z}_c = \widehat{j}_c(X) + S_e\{\Delta_e - \widehat{j}_c(X)\}/\widehat{q}_c(W),$$

combining the actual-change imputation $\widehat{j}_c$ of the previous section (observed pairs, fixed unit weights, `imputation_pooling`) with the fitted observation probability $\widehat{q}_c$ of the conditional-MTAR section (`selection_model`, `selection_pooling`, fitted on every F row of the cohort). Δ_e and $1/\widehat{q}_c$ are evaluated only on observed pairs ($S_e = 1$); the residual term is not renormalized, $\widehat{q}$ enters only through the floored inverse $1/\max(\widehat{q}, 10^{-6})$ (see Probability clipping), and no fill-only rule is used. The repeat rates a_k, synthetic unit masses wa_k, and normalizers retain the pure-imputation conventions. Treatment IPW uses risk-set treatment propensity logits and no counterfactual regression; OR uses a counterfactual control-mean regression fitted to $\widehat{Z}$ with the requested `outcome_pooling`, and no treatment propensity; DR uses both. It is one estimator, not an average of the weighted and imputed estimates, and it has no general finite equality with either.

Two formulas, two designs: `xformla` builds the imputation, counterfactual regression and treatment designs, evaluated on every F row; `selection_xformla` builds the observation-probability design only (`NULL` inherits `xformla`), also on every F row. A variable that appears only in `selection_xformla` never enters the outcome or treatment models. Known cohorts and complete covariates of both formulas are required on every F row, including frame-only, singleton and zero-observation units. The three pooling arguments are independent: `selection_pooling` and `imputation_pooling` each offer separate, within-cohort and across-cohort slopes; `outcome_pooling` offers separate or pooled

counterfactual slopes. Pair and date observation models, intercept-only and no-intercept formulas retain their existing alias and rank policies.

Every active fit must meet its existing regularity, even where a correctness branch below would make it redundant: a cell whose pairs (pair model) or dates (date model) are all observed or all unobserved, with its own free intercept, is refused through the existing separation/boundary refusals, although pure imputation would run on the same data. Imputation, treatment-logit and counterfactual-regression failures also refuse. Nothing falls back to pure imputation or to probability weighting.

Sufficient (not necessary) conditions for each primitive link: conditional mean trend-MAR for every needed cell given the common conditioning set W (the `xformla` covariates and any variables only in `selection_xformla`), with positive observation probability; for each needed cell EITHER the imputation arm (a correct linear actual-change mean given W, represented by the `xformla` design, for every cell of its imputation pooling group) OR the observation arm (a correct observation model given W for every cell of its selection block; for `"independent_period_logit"` correct date logits at both endpoint blocks and conditional independence of the two used dates given cohort and W); the arms may differ across cells; and in addition a correct synthetic treatment propensity logit for treatment IPW, a correct counterfactual control-mean regression for treatment OR, or either for treatment DR. Regression correctness concerns the relevant block (all shared blocks when `outcome_pooling="pooled"`), each of whose cells needs one valid reconstruction arm, with PT/no anticipation and the fixed-weight population law. Observation-layer robustness does not make treatment IPW or OR doubly robust in the treatment layer. Shared coefficients and the two endpoints of a date model couple cells, so a reconstruction error elsewhere in a shared group can affect its other predictions. Correctness of an `xformla`-only imputer alone does not establish reconstruction when variables only in `selection_xformla` shift the outcome-change mean; a correct observation model given W does. Adding selection covariates is therefore not automatically harmless. When both reconstruction arms of a needed cell are wrong, or the selected treatment adjustment lacks its required correct model, the link is generally biased. Recovered joint ATT/GMM coordinates additionally need valid catalog moments for every used link and the accepted identification premises; with incompatible per-link pseudo-targets the limit depends on the weighting. There is no finite-sample unbiasedness; small fitted probabilities inflate the variance.

Complete original-unit influence functions on all F rows carry the observation-probability, imputation, repeat-rate, treatment-propensity and counterfactual-regression paths when active, and both normalizers into the existing GMM, aggregation, clustering, multiplier-band and pretest consumers with core scaling $\psi = (n_A/N)\phi_F$. `chained$selection` reports `method = "augmented"`, both method axes, `selection_model`, `selection_pooling`, `imputation_pooling`, the selection formula, the assumption and inference text and every fit count. The `selection_formula_inherited` field flags the `NULL` fallback; `chained$nuisance_diagnostics` reports the cells (unit counts, probability-weighted and discrepancy masses, imputed and augmented totals), the imputation groups, the observation-probability blocks (`records$selection$blocks`), the counterfactual regression on $\widehat{Z}$, the treatment logits, the fitted catalog and study-population support; `chained$outcome_design` records both designs. B/A fitting samples, transport columns, custom IPW callbacks, SMAR/Corollary 2, singleton outcome levels, cross-fitting/ML nuisances and efficiency claims are outside this route. No numerical certificate or coverage claim is made.

Probability clipping (chained mode)

Package-owned chained inverse-probability weights clip fitted probabilities with the epsilon 10^{-6} of did 2.5.1 and DRDID 1.3.0. Every risk-set treatment propensity is capped above, $p_c = \min(p, 1 -$

10^{-6}), and the consumed control odds is $p_c/(1-p_c)$ (at most about 10^6). The cap is one-sided, and the upstream control trimming (`trim.level = 0.995`) is not applied: no row, unit or comparison is dropped. Every inverse observation denominator is floored at 10^{-6}: a fitted direct pair probability p_U, the final product $p_s p_t$ of the two endpoint date probabilities (the endpoints themselves are not floored) and a strictly positive empirical pair observation rate. The multiplier $a_U/(\rho \max(p_U, 10^{-6}))$ keeps the empirical repeat rate a_U and the known ρ, so it is bounded by $a_U/(\rho 10^{-6})$ rather than by 10^6. The augmented correction uses the same floored inverse. Pure imputation has no observation denominator; its treatment odds are capped. Outcome and imputation predictions are not probabilities and are not clipped. No argument controls this; ordinary and repeated cross-section mode delegate to did unchanged.

Every logit score and information matrix uses raw probabilities. Observation weights in corrected treatment and outcome fits use the floored denominators; the raw empirical rates and their influence functions are kept. Rank, QR, separation, convergence, boundary and nonfinite refusals remain, as do zero-support, invalid-rate and nonpositive normalizer refusals; clipping does not rescue a failed fit. Finite predictions replayed on other rows are evaluated stably and clipped rather than overflowing. Influence functions differentiate the clipped weights: a capped odds or a floored denominator has no derivative in its probability coefficients away from the cutoff. A value exactly at a cutoff is a kink; the strict activity convention assigns it a zero probability derivative, not an ordinary smooth derivative. The repeat-rate numerator, every normalizer and all other channels are retained.

Ordinary influence-function inference for the clipped estimator also needs a negligible threshold-crossing remainder, for example no probability mass of the fitted probabilities at either cutoff. When clipping binds with positive population probability, the estimand is the clipped functional, which can differ from the original ATT target and can change the usual observation or treatment robustness branches. Original-target claims require clipping to be inactive in population or additional conditions preserving that target; no coverage of the original target follows from software checks. The GMM argument that omits estimated-weight derivatives assumes compatible comparison moments; clipped moments that are incompatible remove that premise. Every fitted treatment model, selection-weighted outcome-regression support and augmented cell records a compact summary of its raw and clipped probability or rate ranges and the numbers of capped or floored rows among the route diagnostics. Treatment summaries include the largest fitting weight; selection-weighted outcome-regression supports include the largest WLS weight, and augmented cells include the largest observed correction weight $wa_U/\max(q, 10^{-6})$.

References

Bellego, C., Benatia, D. and Dortet-Bernadet, V. (2025). The Chained Difference-in-Differences. *Journal of Econometrics* 248, 105783. doi:10.1016/j.jeconom.2024.105783

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics* 225(2), 200–230. doi:10.1016/j.jeconom.2020.12.001

See Also

`aggte`, `conditional_did_pretest`, `tidy`, `glance`, `ggdid`, `vignette("cdid", package = "cdid")`

Examples

```
mpdta <- did::mpdta
```

```
set.seed(1)
out <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                    gname = "first.treat", xformla = ~lpop, data = mpdta,
                    mode = "ordinary", bstrap = FALSE)
summary(out)
cdid::aggte(out, type = "dynamic")

# For panel data the default mode is "chained" and default est_method is "dr".
# Unsupported est_method strings are refused rather than replaced by IPW:
res <- try(cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                        gname = "first.treat", data = mpdta,
                        est_method = "stdipw"), silent = TRUE)
cat(res)


# Chained doubly robust (default est_method = "dr") and IPW-GMM:
ch_dr <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                      gname = "first.treat", xformla = ~lpop, data = mpdta,
                      bstrap = FALSE, cband = FALSE)
summary(ch_dr)
cdid::aggte(ch_dr, type = "dynamic")

# Chained IPW-GMM with analytic inference and the default controls (no user
# estimated-work cap; 1 GiB counted numeric workspace; one worker).
ch <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                   gname = "first.treat", xformla = ~lpop, data = mpdta,
                   est_method = "ipw", gmm_weighting = "optimal",
                   gmm_covariance = "tikhonov", bstrap = FALSE, cband = FALSE)
summary(ch)
cdid::aggte(ch, type = "dynamic")   # no refit

# OPTIONAL finite estimated-work cap (refuses before a stage whose conservative
# work estimate exceeds it; not a CPU/time guarantee) and the existing
# configurable counted numeric workspace (not process RSS):
ch_capped <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                          gname = "first.treat", xformla = ~lpop, data = mpdta,
                          est_method = "ipw", bstrap = FALSE, cband = FALSE,
                          control = list(counted_work_limit = 1e12,
                                         numeric_workspace_bytes = 1073741824))

# Explicit identity weighting (two_step is an alias for optimal):
ch_id <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                      gname = "first.treat", xformla = ~lpop, data = mpdta,
                      est_method = "ipw", gmm_weighting = "identity",
                      bstrap = FALSE, cband = FALSE)

# Pure actual-change imputation as the observation correction (no
# observation-probability model; selection_model etc. are inactive):
ch_imp <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                       gname = "first.treat", xformla = ~lpop, data = mpdta,
                       est_method = "dr", selection = "conditional_mtar",
                       selection_est_method = "imputation",
                       imputation_pooling = "separate",
```

```
                        bstrap = FALSE, cband = FALSE)
ch_imp$chained$selection$imputation_pooling

# Augmented observation correction (selection_est_method = "dr" with
# est_method = "dr"): impute the actual pair changes and correct the observed
# residuals by fitted observation probabilities. Its observation logits need
# every needed pair to be partially observed, so this illustration removes a
# deterministic 1-in-7 pattern of county-years (mpdta itself is balanced).
mp_obs <- mpdta[(mpdta$countyreal + mpdta$year) %% 7 != 0, ]
ch_aug <- cdid::att_gt(yname = "lemp", tname = "year", idname = "countyreal",
                        gname = "first.treat", xformla = ~lpop, data = mp_obs,
                        est_method = "dr", selection = "conditional_mtar",
                        selection_est_method = "dr",
                        imputation_pooling = "within_cohort",
                        selection_model = "pair_logit",
                        selection_pooling = "within_cohort",
                        bstrap = FALSE, cband = FALSE)
ch_aug$chained$selection[c("method", "selection_model", "selection_pooling",
                           "imputation_pooling", "selection_formula_inherited")]
```

`conditional_did_pretest` *Conditional pretrend test (chained by default for panels)*

Description

`conditional_did_pretest()` keeps the 19 formals of `did::conditional_did_pretest()` (did 2.5.1) in order and appends the existing chained reference and selection settings it needs. For panel data the default `mode = NULL` runs the chained conditional pretrend test, matching `att_gt()`; `mode = "ordinary"` (and repeated cross sections) forwards the upstream arguments to `did::conditional_did_pretest()` unchanged, with the `NULL` unbalanced-panel sentinel mapped to did's `FALSE`. The appended chained arguments stay lazy and are neither evaluated nor forwarded in ordinary mode. As in did, the default `est_method` is `"ipw"`, independently of `att_gt()`'s `"dr"` default.

Usage

```
conditional_did_pretest(
  yname,
  tname,
  idname = NULL,
  gname,
  xformla = NULL,
  data,
  panel = TRUE,
  allow_unbalanced_panel = NULL,
  control_group = c("nevertreated", "notyettreated"),
  weightsname = NULL,
```

```
  alp = 0.05,
  bstrap = TRUE,
  cband = TRUE,
  biters = 1000,
  clustervars = NULL,
  est_method = "ipw",
  print_details = FALSE,
  pl = FALSE,
  cores = 1,
  anticipation = 0,
  base_period = "varying",
  fix_weights = NULL,
  mode = NULL,
  frame = NULL,
  selection = "none",
  selection_model = "pair_logit",
  selection_pooling = "separate",
  selection_xformla = NULL,
  selection_universe = "F",
  selection_rho = NULL,
  selection_justification = NULL,
  selection_est_method = "ipw",
  imputation_pooling = "separate",
  control = NULL,
  outcome_pooling = "separate"
)

## S3 method for class 'cdid_MP.TEST'
print(x, ...)

## S3 method for class 'cdid_MP.TEST'
summary(object, ...)
```

Arguments

`yname`, `tname`, `idname`, `gname`, `xformla`, `data`, `panel`, `control_group`, `weightsname`, `clustervars`, `print_details`
: As in `did::conditional_did_pretest()` and `att_gt()`. In chained mode the test coordinates are the `model.matrix(xformla)` columns; variables that appear only in `selection_xformla` are not test coordinates.

`allow_unbalanced_panel`
: `NULL` (default) keeps every eligible unit in chained mode (as `att_gt()`); `FALSE` is refused in chained mode. In ordinary mode `NULL` means did's `FALSE`.

`alp`
: Significance level of the reported critical value.

`bstrap`, `cband`
: Validated only. As in did, the conditional test always uses multiplier draws and reports no standard errors.

`biters`
: Number of multiplier draws.

`est_method` `"ipw"` (default), `"reg"` or `"dr"` in chained mode. A custom function returns an explicit not-applicable result: no conditional process is claimed for arbitrary callbacks.

`pl`, `cores` Multiplier draws may use `cores` worker processes when `pl = TRUE` and `control$workers` declares them; signs are drawn in the calling session, so results equal `cores = 1`.

`anticipation`, `base_period`, `fix_weights`, `mode`, `frame`, `selection`, `selection_model`, `selection_pooling`, `selection_xformla`, `selection_universe`, `selection_rho`, `selection_justification`, `selection_est_method`, `imputation_pooling`, `outcome_pooling`, `control`
The existing chained settings of `att_gt()` with the same meanings; they select the fitted route whose auxiliary comparisons are tested. GMM weighting and PT-all settings are not test arguments.

`x`, `object` A `cdid_MP.TEST` result.

`...` Unused.

Details

Null: for every declared pre-treatment cell c of the selected `base_period` family (the pretrend test cells of `att_gt()`, reference zeros excluded), $E_g[\delta_c \mid z_c] = 0$, where z_c is the `model.matrix(xformla)` row of the unit at the cell's own reference date (the earlier endpoint of its auxiliary comparison). This can be weaker than conditioning on all scientific variables. The process $J_c(u) = \sigma_c E_g[1\{z_c \leq u\}\delta_c]$ is the accepted auxiliary link of the fitted union catalog with the grid indicator inserted into its numerator integrands: IPW uses the outcome change, OR/DR the link residual and pure imputation/augmented correction the imputed or augmented response. The nuisance fits (rates, selection, propensity, imputation and the requested separate or pooled outcome regressions) are reused unchanged; only their loadings depend on u, and the complete original-unit influence function includes all of their forcing (not h_u times the scalar influence function). The links use the probability clipping of `att_gt()` (section Probability clipping): capped treatment odds and floored observation denominators enter the term weights, and the loading integrands carry the same derivative masks. When clipping binds with positive population probability, the tested null concerns the clipped auxiliary functional.

Grid: one row per supplied original unit (frame order; singletons, units without observed outcomes and frame-only units included; duplicate rows keep their mass), evaluated through the route's frozen terms, factor levels, contrasts and transform state; grid-only rows never refit them. Dated covariates are read at each cell's recipe date. A required grid covariate that is missing, or a level absent from the frozen design, makes the complete test unavailable with the unit ids (no unit is deleted and no available-only subset is tested); estimated ATTs are unaffected. The statistic is the unstandardized $CvM = (N/M)\sum_c \sum_m J_c(u_m)^2$ with $M = N$; multiplier draws use one Rademacher sign matrix shared by all cells and grid points, drawn from the caller's RNG stream in the existing batch order, at the unit level or over the full original-unit cluster alignment (clusters without influence included). The p-value is the share of draws exceeding CvM; the critical value is their type-1 quantile. A globally degenerate process (every cluster process row exactly zero) is unavailable. Grid, indicator, process and multiplier work is streamed in bounded chunks and charged against the existing resource caps; no original-unit by grid influence array is formed.

Calibration requires uniform process and multiplier-bootstrap regularity conditions, including a growing number of independent clusters for clustered inference; scalar ATT asymptotic normality alone does not establish them, and finite checks establish algebra, not size or power. This is not the pretrend Wald test: the unconditional `W`/`Wpval` of `att_gt()` is separate and unchanged.

Value

In chained mode an object of class `c("cdid_MP.TEST", "MP.TEST")` with the familiar fields `CvM`, `CvMb`, `CvMcval`, `CvMpval`, `clustervars` and `xformla`; the KS fields are `NULL`. The `chained` element records the null, status and reason, base period, tested cells (with their scalar values), route, grid rule and columns, statistic and multiplier conventions, clusters, RNG provenance, execution and resource accounting. Unavailable tests have status `"unavailable"` (or `"not_applicable"` for custom estimators) and `NA` statistics. In ordinary mode did's `MP.TEST` object.

See Also

`att_gt`, `aggte`, `reexports`, `vignette("cdid", package = "cdid")`

Examples

```
## Not run:
pre <- cdid::conditional_did_pretest(yname = "y", tname = "period", idname = "id",
  gname = "G", xformla = ~x, data = panel_data, control_group = "notyettreated",
  biters = 199)
summary(pre)
pre$chained$cells

## End(Not run)
```

`reexports` *Functions re-exported from did*

Description

These did functions are re-exported unchanged so that the usual workflow works with `cdid::` qualification. See their documentation in the did package: `did::ggdid`, `did::tidy` and `did::glance` (the `generics` methods for `MP` and `AGGTEobj` objects). For chained results (`cdid_MP` and `cdid_AGGTEobj`), **cdid** registers S3 methods for `print()`, `summary()`, `tidy()`, `glance()` and `ggdid()` that extend upstream output with chained support and inference metadata. `conditional_did_pretest()` is cdid's own wrapper (see `conditional_did_pretest`): chained by default for panels, and **did**'s `conditional_did_pretest()` unchanged in ordinary mode.

Details

For chained results:

- `print(x)` calls `summary(x)`. Both print the estimate table with chained support, parallel-trends and inference information, then return the original object invisibly. For `cdid_MP`, the summary also reports unit counts, observation correction, GMM weighting and pretrend test status.
- `tidy(x)` returns one row per reported cell (`cdid_MP`) or aggregate bin (`cdid_AGGTEobj`); `type = "group"` also includes an overall row. It keeps **did**'s columns (`estimate`, `std.error`, `conf.low`, `conf.high`, `point.conf.low`, `point.conf.high`), updating pointwise `statistic`

and p.value, and adding target.id, role ("pre", "post", "reference"), available, support.status, support.reason, certified and band.family for cdid_MP, or label, available, restricted, n.intended, n.used, excluded.ids, reason, certified and band.family for cdid_AGGTEobj.

- glance(x) returns a one-row data frame adding estimator, n.requested, n.available, band.family, inference and clustered for cdid_MP, or estimator, status, na.rm, n.bins, n.bins.available, overall.status, band.family and inference for cdid_AGGTEobj.
- ggdid(x, ...) returns a **ggplot2** plot of cdid_MP cells or cdid_AGGTEobj "group", "dynamic" and "calendar" aggregates ("simple" has no plot) and attaches the plotted grid or bin status table in attr(p, "cdid_presentation_grid") or attr(p, "cdid_presentation_bins").

See Also

att_gt, aggte, conditional_did_pretest, vignette("cdid", package = "cdid")

Index

aggte, 3, *3*, *11, 12*, *18*, *23, 24*
att_gt, *3, 4*, 5, *23, 24*

cdid *(*cdid-package*)*, 2
cdid-package, 2
conditional_did_pretest, *3, 4*, *10*, *18*, 20, *23, 24*

ggdid, *3, 4*, *11*, *18*
ggdid *(*reexports*)*, 23
glance, *3, 4*, *11*, *18*
glance *(*reexports*)*, 23

print.cdid_MP.TEST *(*conditional_did_pretest*)*, 20

reexports, *3*, *11*, 23, *23*

summary.cdid_MP.TEST *(*conditional_did_pretest*)*, 20

tidy, *3, 4*, *11*, *18*
tidy *(*reexports*)*, 23