Package 'pathintdid'

August 8, 2026

Title Path-Integrated Difference-in-Differences

Version 0.1.0

Description Implements the Path-Integrated Difference-in-Differences ('PI-DiD') framework of Salavi (2026), which treats the treatment effect as a trajectory tau(t) = c1(t) - c0(t) and integrates the baseline-differenced gap over a post-treatment window to obtain a cumulative causal effect and a path-integrated average treatment effect on the treated, together with cluster-robust standard errors, confidence intervals, diagnostic plots, and the pre-treatment parallel-trends, grid-density, and anticipation-robustness checks of section 5 of the companion paper. This approach avoids the endpoint-subtraction bias that arises whenever a transitory policy's effect has fully decayed by the evaluation date, in which case the conventional static difference-in-differences estimate can be zero even though the cumulative benefit delivered to treated units was strictly positive. Formerly distributed as a Stata package under the names 'pidid' and 'pididplot'.

License MIT + file LICENSE

URL https://github.com/FabriceSALAVI/PathIntDID

BugReports https://github.com/FabriceSALAVI/PathIntDID/issues

Depends R (>= 3.5)

Imports graphics, grDevices, stats

Suggests testthat (>= 3.0.0)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

Config/testthat/edition 3

NeedsCompilation no

Author Comlan Apollinaire-Fabrice Salavi [aut, cre]

Maintainer Comlan Apollinaire-Fabrice Salavi <fsalavi@africanschoolofeconomics.com>

Repository CRAN

Date/Publication 2026-08-08 11:50:02 UTC

Contents

`pathintdid` *Path-Integrated Difference-in-Differences (PI-DiD) estimator*

Description

Implements the Path-Integrated Difference-in-Differences framework of Salavi (2026). Instead of comparing treated and control outcomes at a single endline date, the treatment effect is treated as a trajectory $\tau(t) = c_1(t) - c_0(t)$, and the baseline-differenced gap $\hat{\tau}(t) = \tau(t) - \tau(t_0)$ is integrated over the post-treatment window $[t_0, t_1]$ (trapezoidal rule) to obtain the cumulative causal effect

$$\sigma = \int_{t_0}^{t_1} \hat{\tau}(t)\, dt,$$

and the path-integrated average treatment effect on the treated, $\bar{\tau} = \sigma/(t_1 - t_0)$. For comparison, 'pathintdid' also reports the conventional static two-period DiD estimate, $\hat{\tau}(t_1)$, which uses only the two boundary dates and can be zero (or misleadingly small) whenever a transitory intervention's effect has fully decayed by t_1, even though σ remains strictly positive (the "endpoint-subtraction bias").

Usage

```
pathintdid(
  data,
  yname,
  idname,
  tname,
  treatname,
  t0,
  t1 = NULL,
  level = 95,
  se = TRUE,
  subset = NULL
)

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

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

Arguments

`data`	a data frame (or object coercible to one) containing the panel.
`yname`	character scalar; name of the numeric outcome column.
`idname`	character scalar; name of the unit/panel-id column.
`tname`	character scalar; name of the calendar-time/period column.
`treatname`	character scalar; name of the time-invariant 0/1 treatment-group indicator column (1 = treated, 0 = control).
`t0`	numeric scalar; baseline period marking the start of the integration window. Under parallel pre-trends, $\hat{\tau}(t_0) \approx 0$; 'pathintdid' reports it as a diagnostic and, regardless of its value, nets it out of σ, $\bar{\tau}$, and the static estimate by construction.
`t1`	numeric scalar; terminal evaluation date (endline). Defaults to the maximum observed value of 'tname'.
`level`	confidence level, in percent, used for the reported confidence intervals and the critical value in the t-test. Default '95'.
`se`	logical; if 'FALSE', skip the standard-error/t-statistic/ confidence-interval calculation and report point estimates only (faster; also useful if the panel is too small or too unbalanced for inference). Default 'TRUE'.
`subset`	optional logical vector (of length 'nrow(data)'), or an expression evaluated in 'data', restricting the estimation sample – analogous to Stata's 'if'/'in' qualifiers.
`x`	an object of class '"pathintdid"'.
`...`	further arguments (currently unused).
`object`	an object of class '"pathintdid"'.

Details

Unless 'se = FALSE', 'pathintdid' also computes a standard error, t-statistic, p-value, and confidence interval for σ, $\bar{\tau}$, and the static estimate, built from the individual-level, baseline-differenced trajectories $\Delta Y_{i,k} = Y_{i,t_k} - Y_{i,t_0}$ of every unit observed at every date on the integration grid, using a group-specific (treated/control) sample covariance matrix combined as $\hat{\Omega} = \hat{\Sigma}_1/\hat{\pi} + \hat{\Sigma}_0/(1-\hat{\pi})$ and the trapezoidal-rule weight vector w implied by the integration grid, giving $\mathrm{Var}(\hat{\sigma}) = w'\hat{\Omega}w/N$. This requires at least two treated and two control units with a complete record at every grid date within '[t0, t1]'; see Details.

'pathintdid' works with panel or repeated cross-section data for the point estimates ($c_0(t)$, $c_1(t)$, σ, $\bar{\tau}$ are computed from group-time means, so 'id' does not need to be balanced across periods for those). Standard errors, however, require individual-level data and are computed only from units with a complete trajectory across the requested integration grid. If the panel is too unbalanced or too small to support inference, 'pathintdid' returns point estimates only, together with a message explaining why (rather than failing).

See Salavi (2026) for the full identification assumptions (SUTVA, random sampling, path-level parallel trends, no anticipation, path smoothness, moment conditions) under which this standard error is consistent. The companion function [pathintdidrobust()] implements the section 5 robustness checks and specification tests of the companion paper.

Value

An object of class ‘"pathintdid"‘, a list with components:

`sigma` cumulative causal effect.

`att_path` path-integrated ATT, ‘sigma / (t1 - t0)‘.

`did_static` conventional static DiD, ‘tauhat(t1)‘.

`tau_t0` pre-treatment gap at ‘t0‘ (parallel-trends diagnostic).

`t0`, `t1`, `level` the window and confidence level used.

`path` a data frame with columns ‘time‘, ‘c0‘, ‘c1‘, ‘taud‘, giving the group-time paths on the estimation window.

`have_se` logical; were standard errors computed.

`se_sigma`, `t_sigma`, `p_sigma`, `ci_sigma`
standard error, t-stat, p-value, and confidence interval for ‘sigma‘ (when ‘have_se‘).

`se_att`, `t_att`, `p_att`, `ci_att`
ditto for ‘att_path‘.

`se_did`, `t_did`, `p_did`, `ci_did`
ditto for ‘did_static‘.

`N1`, `N0`, `df` number of treated/control units used for inference, and residual degrees of freedom (when ‘have_se‘).

`yname`, `idname`, `tname`, `treatname`, `call`
bookkeeping.

References

Salavi, C. A.-F. (2026). "Path-Integrated Difference-in-Differences (PI-DiD): Identification, Estimation, and Inference for Cumulative Treatment Effects." Working paper, African School of Economics.

See Also

[pathintdidplot()] to visualize the treated/counterfactual paths and the running cumulative effect; [pathintdidrobust()] for the companion specification tests.

Examples

```
data(trainingpanel)

fit <- pathintdid(trainingpanel, yname = "consumption", idname = "id",
                  tname = "time", treatname = "treat", t0 = 0, t1 = 5)
print(fit)

## point estimates only (faster; no inference)
pathintdid(trainingpanel, "consumption", "id", "time", "treat",
           t0 = 0, t1 = 5, se = FALSE)
```

pathintdidplot *Plot the causal impact of a Path-Integrated DiD analysis*

Description

Companion plotting function to [pathintdid()]. Draws two panels: (1) the treated vs. counterfactual paths, with the region between them over '[t0, t1]' shaded to represent the cumulative effect σ; and (2) the running cumulative effect $\sigma(t)$, which plateaus even after $\tau(t)$ has decayed back to zero – the visual signature of the endpoint-subtraction bias documented in Salavi (2026) – with a shaded confidence band around the final plateau, pulled from [pathintdid()]'s standard error.

Usage

```
pathintdidplot(
  data,
  yname,
  idname,
  tname,
  treatname,
  t0,
  t1 = NULL,
  t2 = NULL,
  level = 95,
  se = TRUE,
  subset = NULL,
  xlab = "Time",
  main = "PI-DiD causal impact",
  col_control = "navy",
  col_treated = "firebrick",
  col_shade = grDevices::adjustcolor("grey70", alpha.f = 0.5)
)
```

Arguments

data	a data frame (or object coercible to one) containing the panel.
yname	character scalar; name of the numeric outcome column.
idname	character scalar; name of the unit/panel-id column.
tname	character scalar; name of the calendar-time/period column.
treatname	character scalar; name of the time-invariant 0/1 treatment-group indicator column (1 = treated, 0 = control).
t0	numeric scalar; baseline period marking the start of the integration window. Under parallel pre-trends, $\hat{\tau}(t_0) \approx 0$; 'pathintdid' reports it as a diagnostic and, regardless of its value, nets it out of σ, $\bar{\tau}$, and the static estimate by construction.
t1	numeric scalar; terminal evaluation date (endline). Defaults to the maximum observed value of 'tname'.

`t2`	optional numeric scalar; an intermediate "rejoining" date, marked with a vertical reference line alongside ‘t0‘ and ‘t1‘. Must satisfy ‘t0 < t2 <= t1‘.
`level`	confidence level, in percent, used for the reported confidence intervals and the critical value in the t-test. Default ‘95‘.
`se`	logical; if ‘FALSE‘, skip the standard-error/t-statistic/ confidence-interval calculation and report point estimates only (faster; also useful if the panel is too small or too unbalanced for inference). Default ‘TRUE‘.
`subset`	optional logical vector (of length ‘nrow(data)‘), or an expression evaluated in ‘data‘, restricting the estimation sample – analogous to Stata’s ‘if‘/‘in‘ qualifiers.
`xlab`	x-axis label. Default ‘"Time"‘.
`main`	overall plot title.
`col_control`, `col_treated`, `col_shade`	colors for the counterfactual path, the treated path, and the shaded cumulative-effect region.

Value

Invisibly, a list with elements ‘sigma_final‘, ‘did_static‘, and (if standard errors were available) ‘se_sigma‘, ‘ci_lb‘, ‘ci_ub‘. Called for its side effect of drawing a two-panel base-graphics plot.

See Also

[pathintdid()]

Examples

```
data(trainingpanel)
pathintdidplot(trainingpanel, yname = "consumption", idname = "id",
               tname = "time", treatname = "treat",
               t0 = 0, t1 = 5, t2 = 5,
               main = "Training program: cumulative consumption impact")
```

`pathintdidrobust` *Robustness checks and specification tests for PI-DiD*

Description

Companion function to [pathintdid()] / [pathintdidplot()], implementing the three diagnostics of section 5 ("Robustness checks and specification tests") of Salavi (2026):

1. **Dynamic placebo test for pre-treatment parallel trends** – a joint Wald test on the vector of pre-treatment gaps, and a "pre-treatment cumulative envelope" (PCE) Z-test on the integrated pre-treatment gap. Requires at least one wave observed before ‘t0‘; otherwise it is skipped with a note.

2. **Sensitivity to grid density and quadrature scheme** – compares the trapezoidal estimate of σ on the full post-treatment grid against a coarsened half-density grid, and against Simpson's 1/3 rule. Requires an even number of equally spaced post-treatment intervals between 't0' and 't1'.
3. **Anticipation-robust bounding estimator** – recomputes σ with the reference baseline shifted back by $0, 1, \ldots,$ 'maxanticip' pre-treatment grid points, reporting a full sensitivity curve with standard errors.

All three diagnostics degrade gracefully – with an explicit message, never silently – when the data cannot support them.

Usage

```
pathintdidrobust(
  data,
  yname,
  idname,
  tname,
  treatname,
  t0,
  t1 = NULL,
  level = 95,
  maxanticip = 0,
  subset = NULL
)

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

Arguments

`data`	a data frame (or object coercible to one) containing the panel.
`yname`	character scalar; name of the numeric outcome column.
`idname`	character scalar; name of the unit/panel-id column.
`tname`	character scalar; name of the calendar-time/period column.
`treatname`	character scalar; name of the time-invariant 0/1 treatment-group indicator column (1 = treated, 0 = control).
`t0`	numeric scalar; baseline period marking the start of the integration window. Under parallel pre-trends, $\hat{\tau}(t_0) \approx 0$; 'pathintdid' reports it as a diagnostic and, regardless of its value, nets it out of σ, $\bar{\tau}$, and the static estimate by construction.
`t1`	numeric scalar; terminal evaluation date (endline). Defaults to the maximum observed value of 'tname'.
`level`	confidence level, in percent, used for the reported confidence intervals and the critical value in the t-test. Default '95'.
`maxanticip`	non-negative integer; maximum number of pre-treatment grid points by which to shift the reference baseline back, for diagnostic (C). If '0' (the default), diagnostic (C) is skipped.

subset	optional logical vector (of length ‘nrow(data)‘), or an expression evaluated in ‘data‘, restricting the estimation sample – analogous to Stata’s ‘if‘/‘in‘ qualifiers.
x	an object of class ‘"pathintdidrobust"‘.
...	further arguments (currently unused).

Value

An object of class ‘"pathintdidrobust"‘, a list with components ‘pretest‘ (diagnostic A, or ‘NULL‘), ‘grid‘ (diagnostic B, or ‘NULL‘), ‘anticipation‘ (diagnostic C, or ‘NULL‘), plus bookkeeping (‘t0‘, ‘t1‘, ‘level‘, ‘call‘).

References

Salavi, C. A.-F. (2026). "Path-Integrated Difference-in-Differences (PI-DiD): Identification, Estimation, and Inference for Cumulative Treatment Effects." Working paper, African School of Economics.

See Also

[pathintdid()], [pathintdidplot()]

Examples

```
data(trainingpanel)

## K = 4 (even) post-treatment intervals -> diagnostic (B) runs
rb <- pathintdidrobust(trainingpanel, yname = "consumption", idname = "id",
                       tname = "time", treatname = "treat",
                       t0 = 0, t1 = 4, maxanticip = 2)
print(rb)
```

---

trainingpanel *Example training-program panel*

Description

A simulated 20-unit individual-level panel (10 treated, 10 control) illustrating the Path-Integrated Difference-in-Differences framework of Salavi (2026): household consumption expenditure observed over 6 waves, with genuine unit-level sampling variation (i.e., not a pair of pre-aggregated cohort means), so that standard errors can be computed.

Usage

```
trainingpanel
```

Format

A data frame with 120 rows and 4 columns:

id integer; household identifier, 1-20.

time integer; period, 0 (baseline) through 5 (endline).

treat integer; time-invariant treatment indicator (1 = treated, 0 = control).

consumption numeric; daily per-capita household consumption expenditure (US dollars).

Source

Simulated data accompanying Salavi, C. A.-F. (2026), "Path-Integrated Difference-in-Differences (PI-DiD): Identification, Estimation, and Inference for Cumulative Treatment Effects," working paper, African School of Economics.

Examples

```
data(trainingpanel)
summary(trainingpanel)
```

Index