Package 'pcdid'

September 18, 2025

Type Package

Title Principal Components Difference-in-Differences

Version 1.0.0

Date 2025-09-13

Maintainer Xiaolei Wang <adamwang15@gmail.com>

Description Implements the Principal Components Difference-in-Differences estimators as described in Chan, M. K., & Kwok, S. S. (2022) <doi:10.1080/07350015.2021.1914636>.

License GPL (>= 3)

Imports stats, sandwich, lmtest

Depends R (>= 3.5)

LazyData true

RoxygenNote 7.3.2

Encoding UTF-8

URL https://github.com/adamwang15/pcdid

BugReports https://github.com/adamwang15/pcdid/issues

Suggests tinytest

NeedsCompilation no

Author Marc Chan [aut] (ORCID: <https://orcid.org/0000-0002-1010-7587>), Xiaolei Wang [aut, cre] (ORCID: <https://orcid.org/0009-0005-6192-9061>)

Repository CRAN

Date/Publication 2025-09-18 08:20:02 UTC

Contents

pcdid *Principal Components Difference-in-Differences*

Description

pcdid first uses a data-driven method (based on principal component analysis) on the control panel to compute factor proxies, which capture the unobserved trends. Then, among treated unit(s), it runs regression(s) using the factor proxies as extra covariates. Analogous to a control function approach, these extra covariates capture the endogeneity arising from potentially unparallel trends.

Usage

```
pcdid(
  formula,
  index,
  data,
  alpha = FALSE,
  fproxy = NULL,
  stationary = FALSE,
  kmax = 10,
  nwlag = round(max(data[[index[2]]])^0.25)
)
```

Arguments

formula	regression specification: depvar ~ treatvar + didvar + indepvar | residvar, where depvar is the dependent variable, treatvar is the binary treatment indicator (1 for treated unit(s) and 0 for control unit(s)), didvar is the interaction term of treatvar and post-treatment time indicator, indepvar is a vector of other independent variables, and residvar is a vector of variables used to compute residuals from control units, if residvar is not specified, indepvar will be used
index	vector of length 2 indicating c(id, time)
data	a data frame containing variables to be used
alpha	perform the parallel trend alpha test. (Note: irrelevant if there is only one treated unit.)
fproxy	set number of factors used. If this option is not specified, the number of factors will be automatically determined by the recursive factor number test.
stationary	advanced option: assume all factors are stationary in the recursive factor number test. (Note: irrelevant if fproxy(#) is specified.)
kmax	advanced option: set maximum number of factors in the recursive factor number test; default is 10. (Note: irrelevant if fproxy(#) is specified.)
nwlag	set maximum lag order of autocorrelation in computing Newey-West standard errors; default is int(T^0.25). (Note: irrelevant if there is more than one treated unit.)

Value

A list of class pcdid, the output list includes element:

mg mean-group estimate of the treatment effect

alpha alpha test result

treated list of treated unit regression results

control list of control unit regression results

Author(s)

Xiaolei Wang <adamwang15@gmail.com>

Examples

```
# use all control variables to compute residuals
result <- pcdid(
  lncase ~ treated + treated_post +
    afdcben + unemp + empratio + mon_d2 + mon_d3 + mon_d4,
  index = c("state", "trend"),
  data = welfare,
  alpha = TRUE
)
result$mg

# use no control variable to compute residuals
result <- pcdid(
  lncase ~ treated + treated_post +
    afdcben + unemp + empratio + mon_d2 + mon_d3 + mon_d4 | NULL,
  index = c("state", "trend"),
  data = welfare,
  alpha = TRUE
)
result$mg
```

---

welfare *Welfare caseloads data*

Description

A sample dataset to examine the effects of welfare waiver programs on welfare caseloads in the United States.

Usage

```
data(welfare)
```

Format

A data frame

state state name

statenum state id

trend time trend in months (oct1986 = 1, nov1986 = 2, etc.)

treated 1 if the state is treated, 0 otherwise

treated_post 1 if the state is treated and post-intervention, 0 otherwise

lncase Natural log of per-capita welfare caseload

afdcben Maximum combined AFDC/Food Stamps benefits for a family of three (in hundred dollar per month)

unemp unemployment rate

empratio Natural log of employment-to-population ratio

mon_d2 seasonal dummy (apr-jun)

mon_d3 seasonal dummy (jul-sep

mon_d4 seasonal dummy (oct-dec)

caseload welfare caseload

popn population

empratio_raw raw employment-to-population ratio

south 1 if the state is in the south, 0 otherwise

control 1 if the state is a control unit, 0 otherwise

T0 Number of preintervention periods for the state (=117 if control state)

Source

Supplemental material, doi:10.1080/07350015.2021.1914636

References

Chan, M. K., & Kwok, S. S. (2022). The PCDID approach: difference-in-differences when trends are potentially unparallel and stochastic. Journal of Business & Economic Statistics, 40(3), 1216-1233.

Index