Package 'StuteTest'

July 23, 2024

Title Stute (1997) Linearity Test

Version 1.0.2

Maintainer Diego Ciccia <diego.ciccia@kellogg.northwestern.edu>

Description Non-parametric test, originally proposed by Stute (1997) <https://www.jstor.org/stable/2242560>, that the expectation of a dependent variable Y given an independent variable D is linear in D.

License MIT + file LICENSE

VignetteBuilder R.rsp

Imports Rcpp (>= 1.0.12), dplyr, plm, rnames, stats

LinkingTo Rcpp, RcppArmadillo

Author Diego Ciccia [aut, cre],
Felix Knau [aut],
Doulo Sow [aut],
Clément de Chaisemartin [aut],
Xavier D'Haultfoeuille [aut]

Encoding UTF-8

RoxygenNote 7.3.2

Suggests R.rsp, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-07-23 00:20:13 UTC

Contents

1

---

stute_test	*Linearity test from Stute (1997)*

Description

Linearity test from Stute (1997)

Usage

```
stute_test(
  df,
  Y,
  D,
  group = NULL,
  time = NULL,
  order = 1,
  seed = NULL,
  brep = 500,
  baseline = NULL
)
```

Arguments

df	(data.frame) A dataframe object.
Y	(char) Outcome variable.
D	(char) Treatment/independent variable.
group	(char) Group variable.
time	(char) Time variable.
order	(numeric) If this option is specified with order = k, the program tests whether the conditional expectation of Y given D is a k-degree polynomial in D. With order = 0, the command tests the hypothesis that the conditional mean of Y given D is constant.
seed	(numeric) This option allows to specify the seed for the wild bootstrap routine.
brep	(numeeric) This option allows to specify the number of wild bootstrap replications. The default is 500.
baseline	(numeric) This option allows to select one of the periods in the data as the baseline or omitted period. For instance, in a dataset with the support of time equal to $(2001, 2002, 2003)$, stute_test(..., baseline = 2001) will test the hypotheses that the expectations of $Y_2002 - Y_2001$ and $Y_2003 - Y_2001$ are linear functions of $D_2002 - D_2001$ and $D_2003 - D_2001$. This option can only be specified in panel mode.

Value

A list with stute_test custom class that includes point estimates and p-values from the test. If the test is performed in panel mode with more than 1 periods, the returned object also includes the point estimate and p-value from a joint test on the sum of the individual test statistics.

Overview

This program implements the non-parametric test that the expectation of Y given D is linear proposed by Stute (1997). In the companion vignette, we sketch the intuition behind the test, as to motivate the use of the package and its options. Please refer to Stute (1997) and Section 3 of de Chaisemartin and D'Haultfoeuille (2024) for further details.

This package allows for two estimation settings:

1. `cross-section`. The test is run using the full dataset, treating each observation as an independent realization of (Y, D).

2. `panel`. The test is run for all values of `time`, using a panel with G groups/units and T periods. In this mode, the test statistics will be computed among observations having the same value of `time`. The program will also return a joint test on the sum of the period-specific estimates. Due to the fact that inference on the joint statistic is performed via the bootstrap distribution of the sum of the test statistics across time periods, this mode requires a strongly balanced panel with no gaps.

References

de Chaisemartin, C, D'Haultfoeuille, X (2024). [Two-way Fixed Effects and Difference-in-Difference Estimators in Heterogeneous Adoption Designs](https://ssrn.com/abstract=4284811).

Stute, W (1997). [Nonparametric model checks for regression](https://www.jstor.org/stable/2242560).

Examples

```
set.seed(0)
GG <- 10; TT <- 5;
data <- as.data.frame(matrix(NA, nrow = GG * TT, ncol = 0))
data$G <- (1:nrow(data) - 1) %% GG + 1
data$T <- floor((1:nrow(data)-1)/GG) + 2000
data <- data[order(data$G, data$T), ]
data$D <- runif(n=nrow(data))
data$Y <- runif(n=nrow(data))
stute_test(df = data, Y = "Y", D = "D")
stute_test(df = data, Y = "Y", D = "D", group = "G", time = "T")
stute_test(df = data, Y = "Y", D = "D", group = "G", time = "T", baseline = 2001)
```

Index