

Package ‘rollmatch’

October 11, 2017

Title Rolling Entry Matching

Version 1.0.0

Date 2017-10-09

Author Rob Chew [aut, cre], Kasey Jones [aut, cre], Mahin Manley [aut], Allison Witman [res], Chris Beadles [res], Yiyao (Echo) Liu [res], Ann Larson [res]

Maintainer Rob Chew <rchew@rti.org>

Description Functions to perform propensity score matching on rolling entry interventions for which a suitable “entry” date is not observed for nonparticipants. For more details, please reference Witman, Beadles, Hoerger, Liu, Kafali, Gandhi, Amico, and Larsen (2016) <<https://academyhealth.confex.com/academyhealth/2016arm/meetingapp.cgi/Paper/9375>>.

License MIT + file LICENSE

URL <https://github.com/RTIInternational/rollmatch>

Depends R (>= 3.0.2)

Imports dplyr (>= 0.5.0), magrittr (>= 1.5.0)

Suggests testthat (>= 1.0.2)

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-10-11 17:35:21 UTC

R topics documented:

rem_synthdata	2
rem_synthdata_small	3
rollmatch	4

Index	6
--------------	----------

rem_synthdata

*Synthetic dataset to illustrate rolling entry***Description**

This dataset represents a synthetic population of individuals who resemble Medicare fee-for-service patients in terms of age, race, spending, inpatient visits, ED visits, chronic conditions, and dual eligibility. The quasi-panel dataset contains multiple observations of non-participants (one for each entry period). Participants enter the data once in the baseline period immediately preceding their unique entry into the intervention. Time-varying covariates (e.g., health conditions, spending, utilization) are dynamic for each entry period's non-participant observations.

Usage

rem_synthdata

Format

A data frame with 254,400 observations and 20 variables:

indiv_id The unique identifier for each individual.

entry_q The period in which the individual enrolled in treatment / entered the intervention.

lq Last baseline quarter before entry into the intervention.

quarter Time variable, indicating the quarter that the variables are measured.

treat Treatment indicator variable (=1 if in treatment group and =0 if in control group).

age The patient's age.

is_black Race indicator variable (=1 if identified as Black, =0 if not).

is_disabled Physical disability indicator variable (=1 if identified as disabled, =0 if not).

is_esrd Disease indicator variable (=1 if identified as having End Stage Renal Disease (ESRD), =0 if not).

is_hispanic Ethnicity indicator variable (=1 if identified as Hispanic, =0 if not).

is_male Gender indicator variable (=1 if identified as Male, =0 if not).

is_white Race indicator variable (=1 if identified as White, =0 if not).

lq_ed Indicates the person had an ED visit during LQ.

lq_ip Indicates the person had an inpatient stay during LQ.

yr_ed2 Count of ED visits during quarters LQ-5 to LQ-1.

yr_ip2 Count of inpatient stays during quarters LQ-4 to LQ-1.

months_dual Number of months of dual Medicare-Medicaid eligibility in the previous year.

chron_num Number of chronic conditions.

qtr_pmt Payments during the quarter.

yr_pmt Payments during the previous 4 quarters.

rem_synthdata_small *Synthetic dataset to illustrate rolling entry (small)*

Description

This dataset represents a synthetic population of individuals who resemble Medicare fee-for-service patients in terms of age, race, spending, inpatient visits, ED visits, chronic conditions, and dual eligibility. The quasi-panel dataset contains multiple observations of non-participants (one for each entry period). Participants enter the data once in the baseline period immediately preceding their unique entry into the intervention. Time-varying covariates (e.g., health conditions, spending, utilization) are dynamic for each entry period's non-participant observations. This is a smaller version of rem_synthadata.

Usage

rem_synthdata_small

Format

A data frame with 12,720 observations and 20 variables:

indiv_id The unique identifier for each individual.

entry_q The period in which the individual enrolled in treatment / entered the intervention.

lq Last baseline quarter before entry into the intervention.

quarter Time variable, indicating the quarter that the variables are measured.

treat Treatment indicator variable (=1 if in treatment group and =0 if in control group).

age The patient's age.

is_black Race indicator variable (=1 if identified as Black, =0 if not).

is_disabled Physical disability indicator variable (=1 if identified as disabled, =0 if not).

is_esrd Disease indicator variable (=1 if identified as having End Stage Renal Disease (ESRD), =0 if not).

is_hispanic Ethnicity indicator variable (=1 if identified as Hispanic, =0 if not).

is_male Gender indicator variable (=1 if identified as Male, =0 if not).

is_white Race indicator variable (=1 if identified as White, =0 if not).

lq_ed Indicates the person had an ED visit during LQ.

lq_ip Indicates the person had an inpatient stay during LQ.

yr_ed2 Count of ED visits during quarters LQ-5 to LQ-1.

yr_ip2 Count of inpatient stays during quarters LQ-4 to LQ-1.

months_dual Number of months of dual Medicare-Medicaid eligibility in the previous year.

chron_num Number of chronic conditions.

qtr_pmt Payments during the quarter.

yr_pmt Payments during the previous 4 quarters.

rollmatch

*Rolling entry matching***Description**

rollmatch is the main function of the package `<rollmatch>`, which implements a comparison group selection methodology for interventions with rolling participant entry over time. A difficulty in evaluating rolling entry interventions is that a suitable "entry" date is not observed for non-participants. This method, called rolling entry matching, assigns potential comparison non-participants multiple counterfactual entry periods which allows for matching of participant and non-participants based on data immediately preceding each participant's specific entry period, rather than using data from a fixed pre-intervention period.

Usage

```
rollmatch(formula, data, tm, entry, id, lookback = 1, caliper = 0,
  weighted_pooled_stdev = FALSE, num_matches = 3, match_on = "logit",
  model_type = "logistic", replacement = TRUE)
```

Arguments

formula	A formula in the form <code>treat ~ x1 + x2 ...</code> where <code>treat</code> is a binary treatment indicator (<code>Treat = 1</code> , <code>Control = 0</code>) and <code>x1</code> and <code>x2</code> are pre-treatment covariates. Both the treatment indicator and pre-treatment covariates must be contained in the input dataset.
data	The input panel dataset.
tm	The time period indicator.
entry	The time period in which the participant enrolled in the intervention (in the same units as the <code>tm</code> variable).
id	The individual id variable.
lookback	The number of time periods to look back before the time period of enrollment (1-10).
caliper	The pre-specified distance within which to allow matching. The caliper width is calculated as the <code>caliper</code> multiplied by the pooled standard deviation of the propensity scores or the logit of the propensity scores - depending on the value of <code>match_on</code> .
weighted_pooled_stdev	Option that allows for weighted pooled standard deviation for caliper matching. <code>FALSE</code> for average pooled standard deviation and <code>TRUE</code> for weighted pooled standard deviation.
num_matches	The number of comparison beneficiary matches to attempt to assign to each treatment beneficiary
match_on	Match on estimated propensity score (" <code>pscore</code> ") or logit of estimated propensity score (" <code>logit</code> ").

model_type	Use logistic regression ("logistic") or "probit" regression ("probit") to estimate the predicted probability of participating in the innovation.
replacement	Assign comparison beneficiaries with replacement (TRUE) or without replacement (FALSE). If replacement is TRUE, then comparison beneficiaries will be allowed to be used with replacement within a single quarter, but will be allowed to match to different treatment beneficiaries across multiple quarters.

Details

Rolling entry matching requires three steps. First, a quasi-panel dataset is constructed containing multiple observations of non-participants (one for each entry period). Participants enter the data once in the baseline period immediately preceding their unique entry into the intervention. Time-varying covariates (e.g., health conditions, spending, utilization) are dynamic for each entry period's non-participant observations. Second, a predicted probability of treatment is obtained for participants and non-participants (e.g., through propensity score matching). Finally, the pool of potential comparisons for each participant is restricted to those that have the same "entry period" into the intervention. The matching algorithm then selects the best matched comparison(s) for each participant from the pool of non-participants with the same entry period.

Value

rollmatch returns an object of class "rollmatch".

An object of class "rollmatch" is a list containing the following components:

call	The original rollmatch call.
model	The output of the model used to estimate the distance measure.
scores	The propensity score and logit of the propensity score.
data	The original dataset with matches, scores, and weights applied.
summary	A basic summary table.
balance	The balancing table.

Examples

```
data(package="rollmatch", "rem_synthdata_small")

formula <- as.formula(treat ~ qtr_pmt + yr_pmt + age)

r_match <- rollmatch(formula, data = rem_synthdata_small, tm = "quarter",
                    entry = "entry_q", id = "indiv_id", caliper = 0.2)

r_match
```

Index

*Topic **datasets**

rem_synthdata, [2](#)

rem_synthdata_small, [3](#)

formula, [4](#)

rem_synthdata, [2](#)

rem_synthdata_small, [3](#)

rollmatch, [4](#)