Package 'rollmatch'

February 12, 2024

Title Rolling Entry Matching

Version 2.0.3

Date 2024-02-11

Author Rob Chew [aut, cre], Kasey Jones [aut, cre], Mahin Manley [aut], Allison Witman [res], Chris Beadles [res], Yiyan 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 et al. (2018) <doi:10.1111/1475-6773.13086>.

License MIT + file LICENSE

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

LazyData true

Depends R (>= 3.0.2)

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

Suggests testthat (>= 1.0.2)

RoxygenNote 7.2.0

NeedsCompilation no

Repository CRAN

Date/Publication 2024-02-12 00:40:02 UTC

R topics documented:

---

reduce_data	*Preprocessing Step to Rolling Entry Matching*

Description

Preprocessing Step to Rolling Entry Matching

Usage

```
reduce_data(data, treat, tm, entry, id, lookback = 1)
```

Arguments

data	Original dataset before reduce_data() was ran.
treat	String for name of treatment variable in data.
tm	String for time period indicator variable name in data.
entry	String for name of time period in which the participant enrolled in the intervention (in the same units as the tm variable).
id	String for individual id variable name in data.
lookback	The number of time periods to look back before the time period of enrollment (1-...).

Value

reduced_data returns a dataset of reduced data ready for propensity scoring and to use in the function score_data()

Examples

```
data(package="rollmatch", "rem_synthdata_small")
reduced_data <- reduce_data(data = rem_synthdata_small, treat = "treat",
                            tm = "quarter", entry = "entry_q",
                            id = "indiv_id", lookback = 1)
reduced_data
```

---

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 last of 3 main functions in the rollmatch package *<rollmatch>* 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(
  scored_data,
  data,
  treat,
  tm,
  entry,
  id,
  vars,
  lookback,
  alpha = 0,
  standard_deviation = "average",
  num_matches = 3,
  replacement = TRUE
)
```

Arguments

scored_data	Output from scored_data() or the output from reduce_data() with propensity scores labeled "score".
data	Original dataset before reduce_data() was ran.
treat	String for name of treatment variable in data.
tm	String for time period indicator variable name in data.
entry	String for name of time period in which the participant enrolled in the intervention (in the same units as the tm variable).
id	String for individual id variable name in data.
vars	Vector of column names used in the propensity score algorithm. This is used when creating the balance table.
lookback	The number of time periods to look back before the time period of enrollment (1-...).

alpha   Part of the pre-specified distance within which to allow matching. The caliper
                         width is calculated as the alpha multiplied by the pooled standard deviation of
                         the propensity scores or the logit of the propensity scores - depending on the
                         value of match_on.

standard_deviation
                         String. 'average' for average pooled standard deviation, 'weighted' for weighted
                         pooled standard deviation, and 'None' to not use a standard deviation multipli-
                         cation. Default is "average".

num_matches   Number of comparison beneficiary matches to attempt to assign to each treat-
                         ment beneficiary. Default is 3.

replacement   Assign comparison beneficiaries with replacement (TRUE) or without replace-
                         ment (FALSE). If replacement is TRUE, then comparison beneficiaries will
                         be allowed to be used with replacement within a single quarter, but will not be
                         allowed to match to different treatment beneficiaries across multiple quarters.
                         Default is TRUE.

Details

Rolling entry matching requires preliminary steps. This package will assist the user in steps 2 and
3. 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. The user of
rollmatch is expected to have already created this quasi-panel dataset. Second, the pool of potential
comparisons for each participant is restricted to those that have the same "entry period" into the
intervention (see function "reduce_data"). Finally, a predicted probability of treatment is obtained
for participants and non-participants (e.g. through propensity score matching). The user can use
function "score_data" to complete this step, or create use their own propensity score calculation.

The final step consists of the matching algorithm. The algorithm selects the best matched compar-
ison(s) for each participant from the pool of non-participants with the same entry period. This is
completed via the function "rollmatch".

Value

rollmatch returns an object of class "rollmatch".

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

model   The output of the model used to estimate the distance measure.

scores   The propensity scores used in the matching algorithm.

data   The original dataset with all matches added.

summary   A basic summary table with counts of matched and unmatched data.

ids_not_matched
                         A vector of the treatment IDs that were not matched.

total_not_matched
                         The number of treatment IDs not matched.

matched_data   R data.frame of matches with scores, matching information, and the weights of
                         the individuals

balance table showing the full treatment, full control, matched treatment, and matched comparison group means and standard deviations for the variables used in the model.

Examples

```
data(package="rollmatch", "rem_synthdata_small")
reduced_data <- reduce_data(data = rem_synthdata_small, treat = "treat",
                            tm = "quarter", entry = "entry_q",
                            id = "indiv_id", lookback = 1)
fm <- as.formula(treat ~ qtr_pmt + yr_pmt + age)
vars <- all.vars(fm)
scored_data <- score_data(reduced_data = reduced_data,
                          model_type = "logistic", match_on = "logit",
                          fm = fm, treat = "treat",
                          tm = "quarter", entry = "entry_q", id = "indiv_id")
output <- rollmatch(scored_data, data=rem_synthdata_small, treat = "treat",
                    tm = "quarter", entry = "entry_q", id = "indiv_id",
                    vars = vars, lookback = 1, alpha = .2,
                    standard_deviation = "average", num_matches = 3,
                    replacement = TRUE)
output
```

---

score_data *Create propensity scores using a logistic or probit regression model*

Description

Create propensity scores using a logistic or probit regression model

Usage

```
score_data(reduced_data, model_type, match_on, fm, treat, tm, entry, id)
```

Arguments

reduced_data   Dataframe of reduced treatment and comparison data. See output of reduce_data().

model_type   Use logistic regression ("logistic") or "probit" regression ("probit") to estimate the predicted probability of participating

match_on   Match on estimated propensity score ("pscore") or logit of estimated propensity score ("logit").

fm   A [formula](#) in the form treat ~ x1 + x2 ... where treat is a binary treatment indicator (Treat = 1, Control = 0) and x1 and x2 are pre-treatment covariates. Both the treatment indicator and pre-treatment covariates must be contained in the input dataset.

treat   String for name of treatment variable in data.

tm	String for time period indicator variable name in data.
entry	String for name of time period in which the participant enrolled in the intervention (in the same units as the tm variable).
id	String for individual id variable name in data.

Value

A copy of reduced_data input with added propensity scores

Examples

```
## Not run:
data(package="rollmatch", "rem_synthdata_small")
fm <- as.formula(treat ~ qtr_pmt + age + is_male + is_white)
reduced_data <- reduce_data(data = rem_synthdata_small, treat = "treat",
                            tm = "quarter", entry = "entry_q",
                            id = "indiv_id", lookback = 1)
scored_data <- score_data(reduced_data = reduced_data,
                          model_type = "logistic", match_on = "logit",
                          fm = fm, treat = "treat", tm = "quarter",
                          entry = "entry_q", id = "indiv_id")
head(scored_data)

## End(Not run)
```

Index