CEM_Equity.R

henrydambanemuya

2019-12-09

```
#!/usr/bin/env Rscript

# Install Packages
# install.packages("cem")
# install.packages("reshape")
# install.packages("lme4")

# Import Packages
library(cem)
```

```
## Loading required package: tcltk
```

```
## Loading required package: lattice
```

```
##
## How to use CEM? Type vignette("cem")
```

```
library(readr)
library(xtable)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(reshape)
```

```
##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:dplyr':
##
##     rename
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:reshape':
##
##     expand
```

```
# Set working directory
setwd("~/Documents/CRII/Seedrs")
# Import Data
equity <- read.csv("./Data/equity_cem.csv")

# Independent Variables (Crowd Features)
equity$NumContributors <- as.numeric(equity$NumContributors)
equity$CovInterEventTime <- as.numeric(equity$CovInterEventTime)
equity$CovContributionAmount <- as.numeric(equity$CovContributionAmount)
equity$TimeToFirstContribution..sec. <- as.numeric(equity$TimeToFirstContribution..sec.)
equity$Duration..days. <- as.numeric(equity$Duration..days.)

# Pre-Treatment Control Variables (Platform Features)
equity$EquityPercent <- as.numeric(equity$EquityPercent)
equity$PreMoneyValuation <- as.numeric(equity$PreMoneyValuation)
equity$EisComplinat <- as.factor(equity$EisComplinat)
equity$SeisCompliant <- as.factor(equity$SeisCompliant)
equity$PassedQuiz <- as.factor(equity$PassedQuiz)
equity$SelfCertification <- as.factor(equity$SelfCertification)
equity$Slug <- as.numeric(equity$Slug)

# Dependent Variable (Class Feature)
equity$Status <- as.numeric(equity$Status)

# Compare class sizes
tr <- which(equity$Status==0)
ct <- which(equity$Status==1)

ntr <- length(tr)
nct <- length(ct)

min(equity$Status)
```

```
## [1] 0
```

```
max(equity$Status)
```

```
## [1] 1
```

```
mean(equity$Status[tr]) - mean(equity$Status[ct])
```

```
## [1] -1
```

```
# Platform variables: pre-treatment covariates (not randomly assigned)
vars.platform <- c("EquityPercent", "PreMoneyValuation", "EisComplinat", "SeisCompliant"
,"PassedQuiz", "Slug", "Status")

# Focus on these pre-treatment covariates
# Compute L1 statistic, as well as several unidimensional measures of imbalance
imbalance(group = equity$Status, data = equity[vars.platform], drop = "Status")
```

```
##
## Multivariate Imbalance Measure: L1=0.569
## Percentage of local common support: LCS=28.8%
##
## Univariate Imbalance Measures:
##
##                     statistic   type         L1          min          25%
## EquityPercent     -0.04209636 (diff) 0.13623529  0.009902951 -0.039611804
## PreMoneyValuation  0.01308178 (diff) 0.07119718  0.000000000  0.001515152
## EisComplinat      18.55764902 (Chi2) 0.15560835           NA           NA
## SeisCompliant     16.59414398 (Chi2) 0.15604828           NA           NA
## PassedQuiz        53.03347280 (Chi2) 0.04801670           NA           NA
## Slug               0.02194750 (diff) 0.03161120 -0.006109980  0.019348269
##                           50%          75%         max
## EquityPercent     -0.028876114 -0.082310358  0.49494949
## PreMoneyValuation  0.003232323  0.006060606  0.49494949
## EisComplinat                NA           NA          NA
## SeisCompliant               NA           NA          NA
## PassedQuiz                  NA           NA          NA
## Slug               0.008825526  0.031568228 -0.00101833
```

```
# Automated Coarsening
mat <- cem(treatment = "Status", data = equity[vars.platform], drop = "Status", eval.imb
alance = TRUE, keep.all = TRUE)
mat # L1 Statistic
```

```
##            G0  G1
## All       479 261
## Matched   362 239
## Unmatched 117  22
##
##
## Multivariate Imbalance Measure: L1=0.355
## Percentage of local common support: LCS=45.5%
##
## Univariate Imbalance Measures:
##
##                        statistic   type           L1         min
## EquityPercent     -7.363408e-03 (diff) 9.551205e-02  0.00000000
## PreMoneyValuation -5.234586e-05 (diff) 1.660855e-02  0.00000000
## EisComplinat       1.428895e+01 (Chi2) 8.326673e-17          NA
## SeisCompliant      1.229764e+01 (Chi2) 2.775558e-17          NA
## PassedQuiz         2.517304e+01 (Chi2) 5.551115e-17          NA
## Slug              -1.916129e-03 (diff) 5.898060e-17 -0.00610998
##                              25%          50%          75%         max
## EquityPercent     -0.0099029511 -0.005941771 -0.012378689  0.00000000
## PreMoneyValuation  0.0002020202  0.000000000  0.001010101  0.03030303
## EisComplinat                NA           NA           NA          NA
## SeisCompliant               NA           NA           NA          NA
## PassedQuiz                  NA           NA           NA          NA
## Slug               0.0027155465 -0.004752206  0.003054990 -0.01289885
```

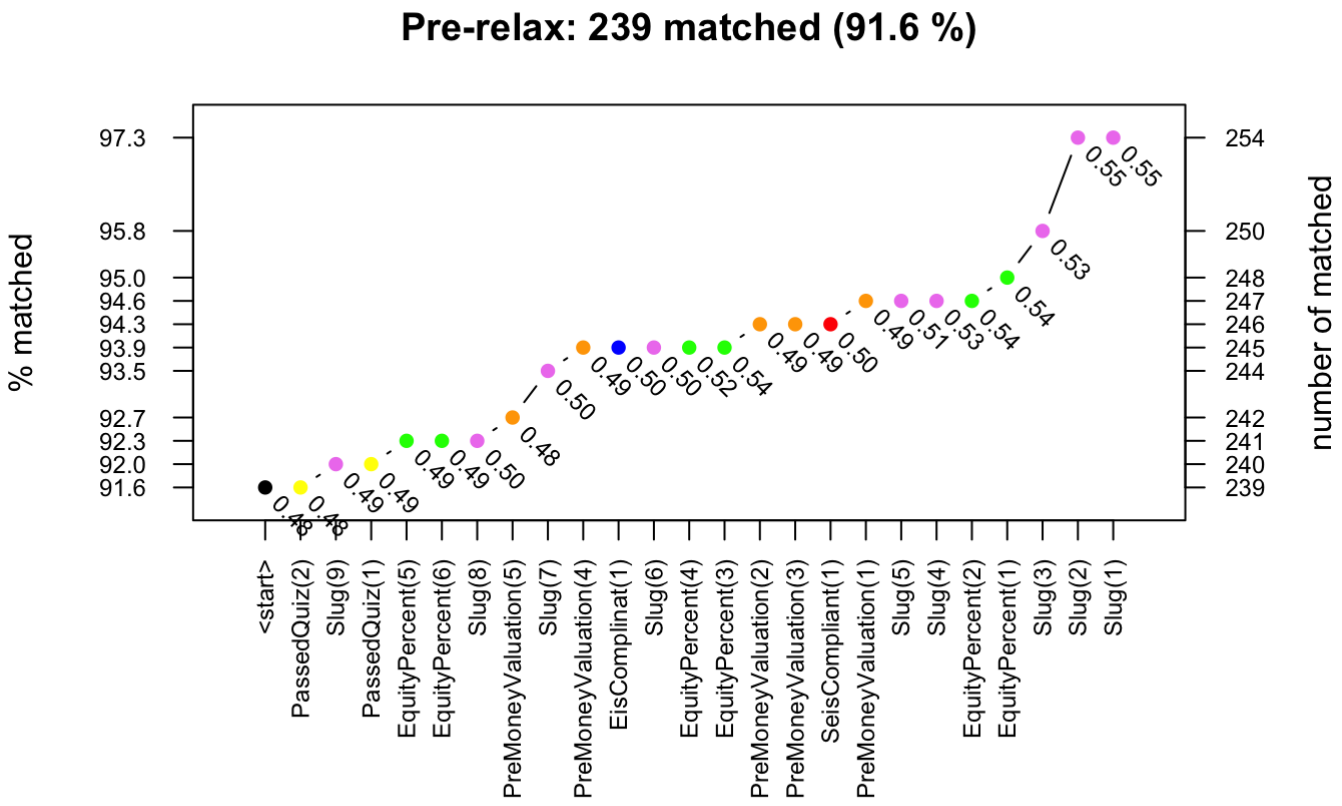
```
# Categorical variables levels
# levels(equity$SeisCompliant)
# levels(equity$EisComplinat)
# levels(equity$PassedQuiz)
# levels(equity$SelfCertification)

# Numerical Variables
# table(equity$EquityPercent)
# table(equity$PreMoneyValuation)
# table(equity$Slug)

#qplot(data = equity, EquityPercent, geom = "histogram", binwidth=1)

# relax matches
tab <- relax.cem(mat, equity, depth = 1, perc = 0.3)
```

```
## Executing 24 different relaxations
##
  |
  |                                                                      |   0%
  |
  |=============                                                         |  20%
  |
  |==========================                                            |  40%
  |
  |========================================                              |  60%
  |
  |=======================================================               |  80%
  |
  |======================================================================| 100%
```




```
## SATT: Estimating the causal effect from cem output

# Appeal
att(mat, NumContributors ~ Status, data = equity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##            G0  G1
## All       479 261
## Matched   362 239
## Unmatched 117  22
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 1.469101 (p.value=0.000340)
## 95% conf. interval: [0.665453, 2.272749]
```

```
# Momentum
att(mat, CovInterEventTime ~ Status, data = equity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##            G0  G1
## All       479 261
## Matched   362 239
## Unmatched 117  22
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.326340 (p.value=0.069800)
## 95% conf. interval: [-0.026414, 0.679094]
```

```
# Variation
att(mat, CovContributionAmount ~ Status, data = equity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##            G0  G1
## All       479 261
## Matched   362 239
## Unmatched 117  22
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.666790 (p.value=0.001269)
## 95% conf. interval: [0.261293, 1.072288]
```

```
# Latency
att(mat, TimeToFirstContribution..sec. ~ Status, data = equity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##            G0  G1
## All       479 261
## Matched   362 239
## Unmatched 117  22
##
## Logistic model on CEM matched data:
##
## SATT point estimate: -0.589324 (p.value=0.002709)
## 95% conf. interval: [-0.974473, -0.204175]
```

```
# Engagement
att(mat, Duration..days. ~ Status, data = equity, model="logit")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```
##
##            G0  G1
## All       479 261
## Matched   362 239
## Unmatched 117  22
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.113854 (p.value=0.558359)
## 95% conf. interval: [-0.267415, 0.495123]
```

```
write.csv(mat$X, file = "./Data/cem_results.csv")
```