CEM_Lending.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/Prosper")
# Import Data
lending <- read.csv("./Data/lending_cem.csv")

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

# Pre-Treatment Control Variables (Platform Features)
lending$AmountRequested <- as.numeric(lending$AmountRequested)
lending$BorrowerRate <- as.numeric(lending$BorrowerRate)
lending$MonthlyLoanPayment <- as.numeric((lending$MonthlyLoanPayment))
lending$ProsperScore <- as.factor(lending$ProsperScore)
lending$CreditGrade <- as.factor(lending$CreditGrade)
lending$DebtToIncomeRatio <- as.numeric(lending$DebtToIncomeRatio)
lending$IsBorrowerHomeowner <- as.factor(lending$IsBorrowerHomeowner)

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

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

min(lending$Status)
```

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

```
max(lending$Status)
```

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

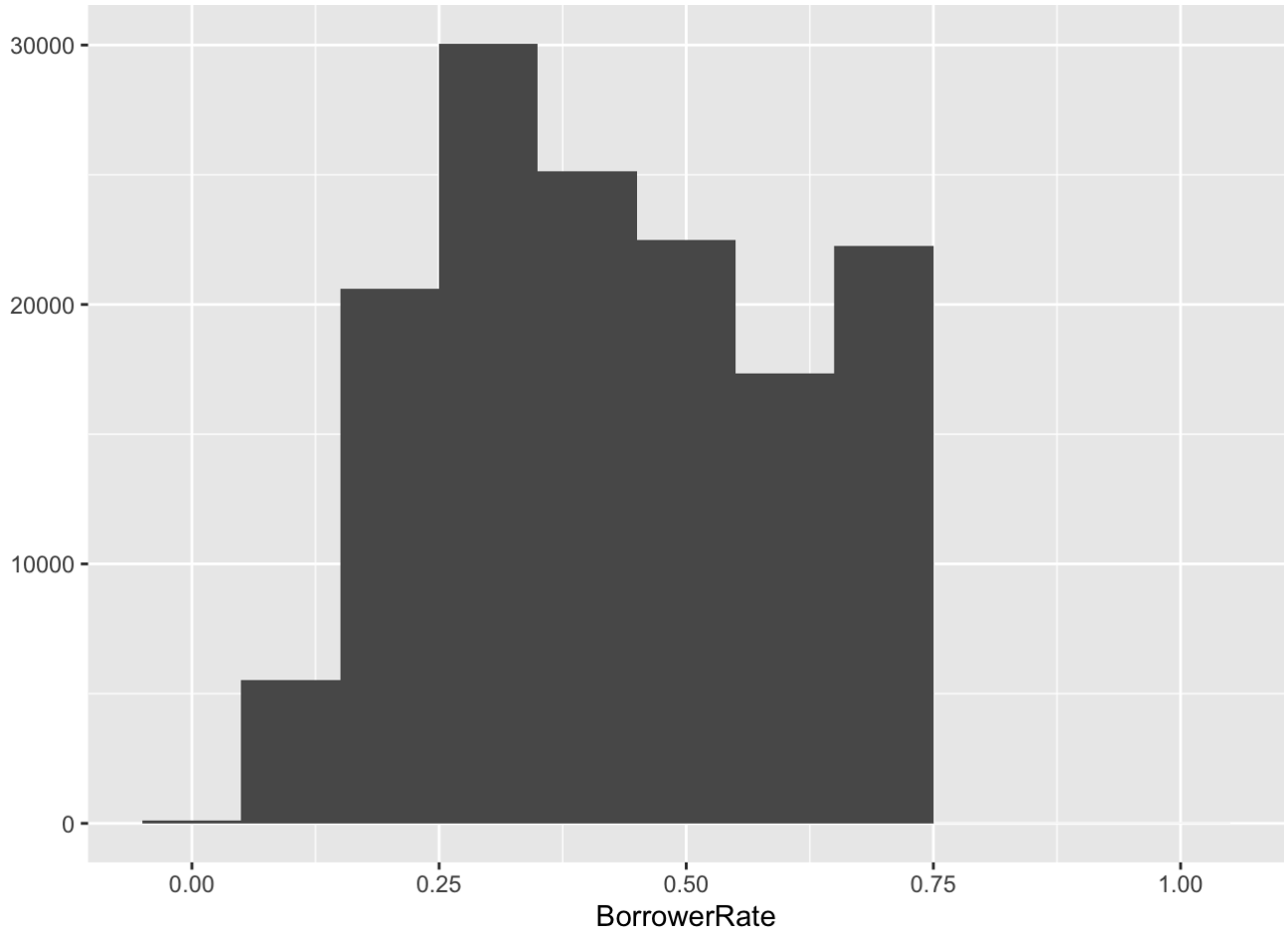
```
ntr <- length(tr)
nct <- length(ct)

# Platform variables: pre-treatment covariates (not randomly assigned)
vars.platform <- c("BorrowerRate", "ProsperScore", "CreditGrade", "DebtToIncomeRatio",
"AmountRequested", "MonthlyLoanPayment", "IsBorrowerHomeowner", 'Status')

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

```
##
## Multivariate Imbalance Measure: L1=0.924
## Percentage of local common support: LCS=3.4%
##
## Univariate Imbalance Measures:
##
##                         statistic   type           L1        min
## BorrowerRate        6.774705e-02 (diff) 0.0001259449 0.00000000
## ProsperScore        9.365032e+04 (Chi2) 0.7383586668         NA
## CreditGrade         1.581979e+04 (Chi2) 0.3418428213         NA
## DebtToIncomeRatio   8.397254e-03 (diff) 0.0461900193 0.00000000
## AmountRequested     6.173362e-02 (diff) 0.0581338551 0.00000000
## MonthlyLoanPayment  6.170580e-02 (diff) 0.0149243443 0.02463525
## IsBorrowerHomeowner 6.277167e+02 (Chi2) 0.0921125965         NA
##                              25%         50%         75%        max
## BorrowerRate        0.0462311558 0.090251256 0.106532663 -0.2763819
## ProsperScore                  NA          NA          NA         NA
## CreditGrade                   NA          NA          NA         NA
## DebtToIncomeRatio   0.0009082652 0.003633061 0.007266122  0.0000000
## AmountRequested     0.0208333333 0.020833333 0.083333333  0.0000000
## MonthlyLoanPayment  0.0270138827 0.041214962 0.082898576  0.0000000
## IsBorrowerHomeowner           NA          NA          NA         NA
```

```
# Automated Coarsening
mat <- cem(treatment = "Status", data = lending[vars.platform], drop = "Status", cutpoin
ts = list(AmountRequested=3), eval.imbalance = TRUE, keep.all = TRUE)
mat # L1 Statistic
```

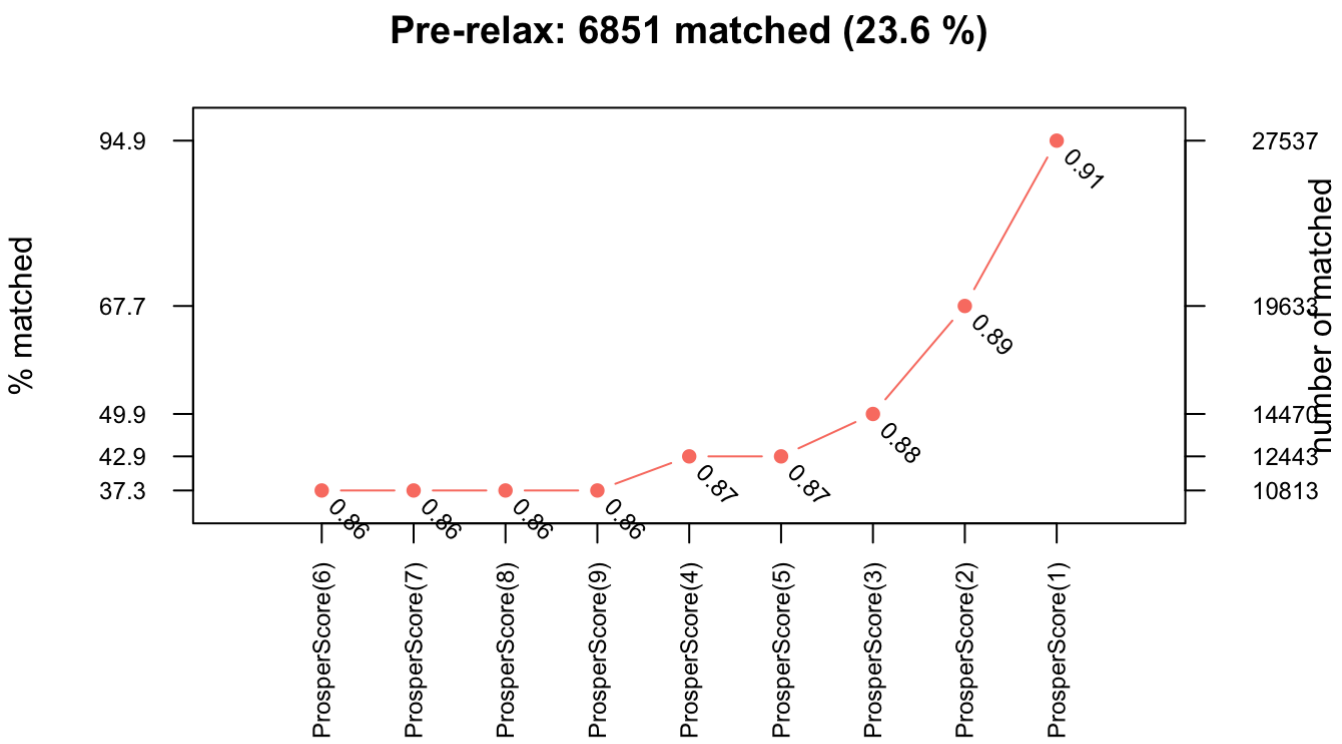
```
##               G0    G1
## All       114536 29013
## Matched    54304  6851
## Unmatched  60232 22162
##
##
## Multivariate Imbalance Measure: L1=0.701
## Percentage of local common support: LCS=10.5%
##
## Univariate Imbalance Measures:
##
##                          statistic   type           L1        min
## BorrowerRate         2.516160e-04 (diff) 2.432735e-05 0.00000000
## ProsperScore         3.682098e+04 (Chi2) 5.551115e-17         NA
## CreditGrade          1.737924e+03 (Chi2) 2.081668e-17         NA
## DebtToIncomeRatio    1.302570e-03 (diff) 5.551115e-17 0.00000000
## AmountRequested     -6.320574e-04 (diff) 1.748277e-02 0.00000000
## MonthlyLoanPayment   2.539305e-03 (diff) 1.576412e-02 0.02532496
## IsBorrowerHomeowner  4.571518e+01 (Chi2) 0.000000e+00         NA
##                              25%           50%           75%         max
## BorrowerRate        0.0036180905 0.0010050251  0.0002010050 0.008040201
## ProsperScore                  NA           NA            NA          NA
## CreditGrade                   NA           NA            NA          NA
## DebtToIncomeRatio   0.0009082652 0.0009082652  0.0002443233 0.000000000
## AmountRequested     0.0000000000 0.0000000000  0.0000000000 0.000000000
## MonthlyLoanPayment  0.0009196215 0.0014590149 -0.0021841012 0.012070033
## IsBorrowerHomeowner           NA           NA            NA          NA
```

```
# Categorical variables levels
# levels(lending$CreditGrade)
# levels(lending$IsBorrowerHomeowner)
# levels(lending$ProsperScore)

# Numerical Variables
# table(lending$AmountRequested)
# table(lending$BorrowerRate)
# table(lending$DebtToIncomeRatio)
# table(lending$MonthlyLoanPayment)

qplot(data = lending, BorrowerRate, geom = "histogram", binwidth=0.1)
```




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

```
## Executing 57 different relaxations
##
  |
  |                                                                      |   0%
  |
  |===========                                                           |  17%
  |
  |======================                                                |  33%
  |
  |==================================                                    |  50%
  |
  |==============================================                        |  67%
  |
  |==========================================================            |  83%
  |
  |======================================================================| 100%
```




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

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

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

```
##
##                G0    G1
## All        114536 29013
## Matched     54304  6851
## Unmatched   60232 22162
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 1.678933 (p.value=0.000000)
## 95% conf. interval: [1.576230, 1.781636]
```

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

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

```
##
##               G0    G1
## All       114536 29013
## Matched    54304  6851
## Unmatched  60232 22162
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 0.708254 (p.value=0.000000)
## 95% conf. interval: [0.638844, 0.777665]
```

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

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

```
##
##               G0    G1
## All       114536 29013
## Matched    54304  6851
## Unmatched  60232 22162
##
## Logistic model on CEM matched data:
##
## SATT point estimate: 1.153931 (p.value=0.000000)
## 95% conf. interval: [1.074709, 1.233154]
```

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

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

```
##
##               G0    G1
## All       114536 29013
## Matched    54304  6851
## Unmatched  60232 22162
##
## Logistic model on CEM matched data:
##
## SATT point estimate: -1.366213 (p.value=0.000000)
## 95% conf. interval: [-1.436401, -1.296025]
```

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

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

```
##
##               G0    G1
## All       114536 29013
## Matched    54304  6851
## Unmatched  60232 22162
##
## Logistic model on CEM matched data:
##
## SATT point estimate: -0.681581 (p.value=0.000000)
## 95% conf. interval: [-0.733888, -0.629274]
```

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