MSD Final Project Report

Forrest Hofmann (fhh2112) & Sagar Lal (sl3946)

2019-05-01 21:53:07

Contents

Introduction

Problem Description

Motivation

Data Source

Reproduction

Reproduction Code

```
teams <- read_csv(here('teams.csv'))

## Parsed with column specification:
## cols(
##   yearID = col_double(),
##   teamID = col_character(),
##   G = col_double(),
##   W = col_double(),
##   L = col_double()
## )

salaries <- read_csv(here('salaries.csv'))

## Parsed with column specification:
## cols(
```

```
##   yearID = col_double(),
##   teamID = col_character(),
##   playerID = col_character(),
##   salary = col_double()
## )
```

```
salaries$salary <- salaries$salary / 1000000
```

```
teams_old <- filter(teams, 1985 <= yearID & yearID <= 1998)
salaries_old <- filter(salaries, 1985 <= yearID & yearID <= 1998)
```

```
teams_old <- teams_old %>%
  mutate(winPercentage = W / (W + L))
```

```
teams_old <- teams_old %>%
  inner_join(salaries_old) %>%
  group_by(yearID, teamID, G, W, L, winPercentage) %>%
  summarize(totalSalary = sum(salary))
```

```
## Joining, by = c("yearID", "teamID")
```

```
salaries_old <- salaries_old %>%
  inner_join(teams_old) %>%
  mutate(salaryShare = salary / totalSalary * 100) %>%
  mutate(salaryShareSquared = salaryShare ^ 2) %>%
  select(yearID, teamID, playerID, salary, salaryShare, salaryShareSquared)
```

```
## Joining, by = c("yearID", "teamID")
```

```
teams_old <- teams_old %>%
  inner_join(salaries_old) %>%
  group_by(yearID, teamID, G, W, L, winPercentage, totalSalary) %>%
  summarize(HHI = sum(salaryShareSquared))
```

```
## Joining, by = c("yearID", "teamID")
```

```
summary(teams_old$winPercentage)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3272  0.4568  0.4984  0.5000  0.5432  0.7037
```

```
summary(teams_old$totalSalary)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.88   12.76   22.32   25.16   36.29   72.36
```

```
summary(teams_old$HHI)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   427.5   668.6   756.3   815.6   879.1  5300.1
```

Reproduction Notes

- original author did not describe how time fixed effects are accounted for (across expansion periods or every year)
- no discussion about limiting to 25 man roster vs 40 man roster
- no discussion of cut players, traded players
- no discussion of signing bonuses

Reproduction Analysis

Extension

Extension Code

```
# TODO: plot distribution of win percentage, total salary, HHI
```

Extension Notes

Extension Analysis