NCAA Football Coaches Salary

Matt Kalin

8/10/2017

Introduction

Payroll plays a large role in every sport. Rich professional teams such as the New York Yankees are perennial championship contenders because of their ability to spend money on talent. Although colleges don't pay their football players millions of dollars, they do spend several figures on head coaches. Some schools also pay millions to their assistant coaching staffs. Schools in the Power 5 conferences (Big Ten, Big 12, ACC, SEC, Pac-12) pay their coaches more money than teams in other conferences (Sun Belt, C-USA, Mountain West, AAC, MAC). The Power 5 conference schools also have the most successful football programs.

Acquire Data

We took salary data from USA Today (http://sports.usatoday.com/ncaa/salaries/). The data are relatively easy to copy from the website and paste into Excel. However, the data is not complete and we need to do additional work to get it in the form we want. The data covers most of the 128 NCAA Division 1 FBS (Football Bowl Subdivision) schools. In order to get the total coaches' payroll, we added the salary of the Head Coach with the Assistant Pay Total. Only 106 of the 128 schools had salary data for both the Head Coach and Assistant Coaches. Penn State, for example, was not included in our dataset because its Assistant Coach salary was not listed. I defined the All.coaches as the team's total coach payroll (the sum of the head coach's total pay and the assistant coach payroll)

```
library(readxl)
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
```

```
NCAAF_coaches_salaries <- suppressWarnings(read_excel(
  "~/Desktop/Summer 2017/NCAAF coaches salaries.xlsx", sheet = "Salary",
  col_types = c("numeric", "text", "text", "text", "numeric", "numeric",
                "numeric", "numeric", "numeric", "numeric")) %>%
  mutate(All.coaches = `TOTAL PAY` + `ASST PAY TOTAL`)) %>%
  filter(!is.na(`ASST PAY TOTAL`)) %>% # filter out teams without asst payroll included
  dplyr::rename(HC.Salary = `TOTAL PAY`, Coach.Payroll = `All.coaches`, Asst.Payroll = `ASST PAY TOTAL`) %>%
  select(SCHOOL, CONF, HC.Salary, Asst.Payroll, Coach.Payroll)
head(NCAAF_coaches_salaries)
```

```
## # A tibble: 6 x 5
##   SCHOOL        CONF    HC.Salary Asst.Payroll Coach.Payroll
##   <chr>         <chr>       <dbl>        <dbl>         <dbl>
## 1 Michigan      Big Ten   9004000      4308750      13312750
## 2 Alabama       SEC       6939395      5320000      12259395
## 3 Ohio State    Big Ten   6094800      4583100      10677900
## 4 Oklahoma      Big 12    5550000      4390900       9940900
## 5 Florida State ACC       5250000      4586000       9836000
## 6 Texas         Big 12    5200130      4546250       9746380
```

Measuring Team Strength

We wanted to compare each school's payroll to its success, so we looked at Jeff Sagarin's College Football ratings at sagarin.com (http://sagarin.com/sports/cfsend.htm). The data could be copied from the website and pasted into Excel, but there were some problems at first. The data contains many columns, but we were only interested in a few: Team, Wins, Losses, and Strength of Schedule (we measured team strength using schedule-adjusted winning percentage, but Sagarin's team power ratings which he provides on his website could also have been used).

The key issue with Sagarin's data is that it copies into Excel as one column. The "Text to Columns" feature (under the Data tab) on Excel can be used to solve this. After some blank or useless columns and rows are deleted, the dataset should be easy to work with.

```
NCAAF_Sagarin <- read_excel("~/Desktop/Summer 2017/NCAAF coaches salaries.xlsx",
                            sheet = "Sagarin")
head(NCAAF_Sagarin)
```

```
## # A tibble: 6 x 6
##   Team       RATING     W     L SCHEDL PREDICTOR
##   <chr>       <dbl> <dbl> <dbl>  <dbl>     <dbl>
## 1 Clemson     105.     14     1   78.2     101.
## 2 Alabama     105.     14     1   80.0     105.
## 3 Michigan     94.0    10     3   72.1      97.4
## 4 Washington   93.3    12     2   71.8      93.3
## 5 Ohio State   93.3    11     2   77.1      96.1
## 6 Oklahoma     93.2    11     2   75.0      90.4
```

I defined two metrics here: schedule rating and win/loss rating. Schedule rating is based on Sagarin's SCHEDL rating but I put it on an ELO rating scale. Sagarin's ratings typically have a mean of 55 and standard deviation of 16 for all division 1 teams. Win/Loss rating uses the ELO inverse formula based on the team's winning percentage that season. Every FBS team won and lost at least one game in 2017 so there were no errors taking the base 10 log of each team's winning percentage. I used an inner join to combine the two datasets.

```
NCAAF_data = NCAAF_Sagarin %>%
  mutate(sched.elo = -400 * log10(1/pnorm(SCHEDL, 55, 16)-1),
         wl.elo = -400 * log10(1/(W / (W + L)) - 1),
         Wins.Rating = sched.elo + wl.elo,
         Record = paste(W, L, sep = "-"),
         Rating.Rk = rank(desc(Wins.Rating))) %>%
  dplyr::rename(SCHOOL = Team) %>%
  inner_join(NCAAF_coaches_salaries) %>%
  select(SCHOOL, CONF, HC.Salary, Asst.Payroll, Coach.Payroll, Wins.Rating, Record, Rating.Rk)
```

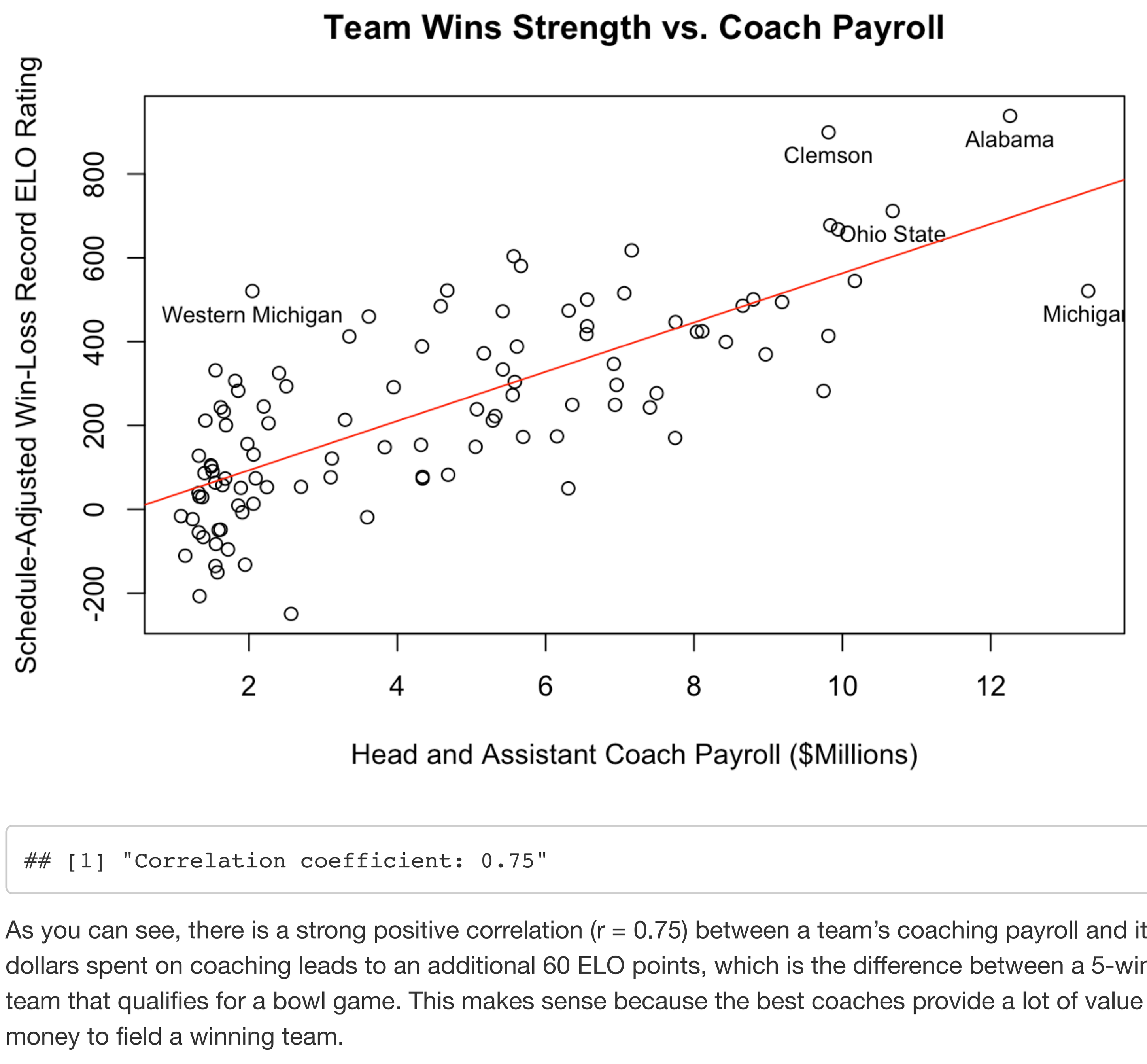
```
## Joining, by = "SCHOOL"
```

```
head(NCAAF_data)
```

```
## # A tibble: 6 x 8
##   SCHOOL CONF  HC.Salary Asst.Payroll Coach.Payroll Wins.Rating Record Rating.Rk
##   <chr>  <chr>     <dbl>        <dbl>         <dbl>       <dbl> <chr>      <dbl>
## 1 Clems… ACC     4422700      5390417       9813117        899. 14-1           2
## 2 Alaba… SEC     6939395      5320000      12259395        939. 14-1           1
## 3 Michi… Big …   9004000      4308750      13312750        521. 10-3          14
## 4 Washi… Pac-…   3605847      3553578       7159425        618. 12-2           6
## 5 Ohio … Big …   6094800      4583100      10677900        712. 11-2           3
## 6 Oklah… Big …   5550000      4390900       9940900        668. 11-2           5
```

Here I plotted each team's total coach payroll against its strength rating.

```
{
  plot((NCAAF_data$Coach.Payroll)/1000000, (NCAAF_data$Wins.Rating), xlab =
         "Head and Assistant Coach Payroll ($Millions)", ylab =
         "Schedule-Adjusted Win-Loss Record ELO Rating",
       main = "Team Wins Strength vs. Coach Payroll")
  plotTeamPoint = function(team){
    team.index = match(team, NCAAF_data$SCHOOL)
    text(NCAAF_data[team.index, "Coach.Payroll"] / 1000000,
         NCAAF_data[team.index, "Wins.Rating"],
         labels = team, pos = 1, cex = 0.8)
  }
  plotTeamPoint("Western Michigan")
  plotTeamPoint("Alabama")
  plotTeamPoint("Clemson")
  plotTeamPoint("Michigan")
  plotTeamPoint("Ohio State")
  {
    ncaaf.payroll.model = lm(Wins.Rating ~ Coach.Payroll, data = NCAAF_data)
    abline(coef(ncaaf.payroll.model)[1], coef(ncaaf.payroll.model)[2] * 1000000, col="red")
  } # trend line
  summary(ncaaf.payroll.model)
  print(paste("Correlation coefficient:", round(cor(NCAAF_data$Wins.Rating, NCAAF_data$Coach.Payroll), 2)))
} # graph
```




```
## [1] "Correlation coefficient: 0.75"
```

As you can see, there is a strong positive correlation (r = 0.75) between a team's coaching payroll and its strength rating. Every additional million dollars spent on coaching leads to an additional 60 ELO points, which is the difference between a 5-win team with no postseason and a 6-win team that qualifies for a bowl game. This makes sense because the best coaches provide a lot of value to the school, so they get paid a lot of money to field a winning team.

I also looked into how the average coach payroll and strength rating differed by conference.

```
confs = c("SEC", "Big Ten", "Big 12", "Pac-12", "ACC", "AAC","Ind.","Mt. West",
        "C-USA", "Sun Belt", "MAC")
NCAAF_Conferences = NCAAF_data %>%
  group_by(CONF) %>%
  dplyr::summarize(Avg.Payroll = mean(Coach.Payroll), Avg.Rating = mean(Wins.Rating)) %>%
  arrange(desc(Avg.Payroll))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
NCAAF_Conferences
```

```
## # A tibble: 11 x 3
##    CONF     Avg.Payroll Avg.Rating
##    <chr>          <dbl>      <dbl>
##  1 SEC         8602234.      436.
##  2 ACC         6929370.      488.
##  3 Big Ten     6731286.      335.
##  4 Big 12      6462033.      378.
##  5 Pac-12      6125829.      338.
##  6 AAC         3921829.      192.
##  7 Mt. West    2384974.      142.
##  8 Ind.        2220890.       33.1
##  9 C-USA       1746600.       44.0
## 10 Sun Belt    1603833.       94.9
## 11 MAC         1442782        67.8
```

Unsurprisingly, the Power Five conferences have the highest paid coaches and the best teams. There is a large drop off between 5th ranked Pac-12 and 6th ranked AAC. This makes sense because the Power Five schools have much more funding to put towards their football program so they can get the best possible coaches and players.