Draft\_Regression

Steven Liao

May 1, 2017

## Salary Analysis

Given the NFL salary cap (currently set for 2017 at 167 million dollars), it was crucial for us to consider salary in our player evaluation analyses. In other words, the value a player brings is not just in how 'good' they are, but also in how 'cheap' they are.

In theory, every general manager strives maximize their return on the salary cap. They can attempt to sign players for cheap (though they are competing against a free market, in theory); however, a more common and - and dependable - method is to build a team using draft picks. Due to a fixed rookie wage scale, drafted players are believed to be paid significantly less than they would be worth on a free market.

Under this framework, we sought to develop a system to fairly value salary, players, and draft picks in conjunction to one another. How much should a player be paid? What is fair compensation for this player (and their contract) in a trade? How much value does a draft pick actually offer, if any?

To start, we needed to develop a metric for the 'fair salary' for a player - this is defined as the 'maximum' salary that a team should be willing to pay for a player in a non edge-case situation (i.e. for instance, one edge case is when a team is far below the minimum spending threshold, and then 'overspending' is preferable to having unspent money go wasted). In other words, by paying players below their fair salary, we are increasing the return on our salary cap!

The premise behind our fair salary metric is that the salary ratio between Player A and Player B should be equal to the talent / contribution ratio. We use Approximate Value (AV) as a proxy for talent and contribution level; Approximate Value is a metric developed by Pro Football Reference designed to measure contribution, and is comparable across positions. It is measured on a yearly basis and is useful as an objective, comparable metric for how much a player contributed to their team in a given year.

In making our fair salary metric, we made the following assumptions:

* AV contribution is linear. In other words, a player whose AV is 8 in a given year is worth twice as much as a player whose AV is 4 in a given year.
* A player with 0 AV is worth $0
* A practice squad player's pay is negligible

Using these three assumptions, and inputting the salary cap for every year from 1994 to 2016 (ranging from 34.6 million in 1994 to 155.27 million in 2016), we were able to calculate the fair salary for each player in that given year, using the following formula:

* Let = total combined salary cap across all 32 teams (167 million times 32 for 2017)
* Let = total number of players who can be on an NFL roster (53 times 32 for 2017)
* Let = Summed AV of the top N NFL players in a given season (averaged per season from 1994 -2016, probably)

Then Fair Salary is:

We began by using the following SQL code to compose a CSV file to be used in R.

sqlite3 -header -csv nfl\_data.db   
"SELECT m.year, player.id, player.name, m.team\_id,   
m.position\_id, m.age, m.av\_value, m.year, m.round,   
m.pick, m.base\_salary, m.cap\_hit, m.dead\_cap   
FROM player,   
((av LEFT OUTER JOIN salary   
ON av.player\_id == salary.player\_id AND av.year == salary.year) k   
LEFT OUTER JOIN draft ON k.player\_id == draft.player\_id) m   
WHERE av.player\_id == player.id   
ORDER BY m.year ASC, m.av\_value DESC" > ..\data\fair\_salary.csv

Here is a view at the data:

## year id name team\_id position\_id age av\_value  
## 1 1994 2403 Steve Young SFO QB 33 23  
## 2 1994 198516 Jerry Rice SFO WR 32 21  
## 3 1994 19893 Barry Sanders DET RB 26 20  
## 4 1994 199145 Ricky Watters SFO RB 25 19  
## 5 1994 199159 Aeneas Williams ARI CB 26 18  
## 6 1994 19912 Eric Turner CLE S 26 18  
## 7 1994 198710 Rod Woodson PIT S 29 18  
## 8 1994 1987150 Greg Lloyd PIT OLB 29 18  
## 9 1994 1985113 Kevin Greene PIT OLB 32 18  
## 10 1994 199017 Emmitt Smith DAL RB 25 17  
## draft\_year round pick base cap\_hit dead\_cap  
## 1 1994 NA NA NA NA NA  
## 2 1994 1 16 2000000 2380000 0  
## 3 1994 1 3 2575000 3080000 0  
## 4 1994 2 45 NA NA NA  
## 5 1994 3 59 750000 1200000 0  
## 6 1994 1 2 1000000 1812500 0  
## 7 1994 1 10 NA NA NA  
## 8 1994 6 150 NA NA NA  
## 9 1994 5 113 NA NA NA  
## 10 1994 1 17 2200000 3200000 0

After manually adding roster size, salary cap size (34.6 million in 1994 to 167 million this year), and number of teams in the league, we were able to use the aforementioned formula to calculate Approximate Value.

Note that in every year in our analyses, AVTotal was simply the summed AV of all players, since the number of players with nonzero AV was always less than the number of rosterable players (N).

Some of our findings included:

* Matt Ryan, the player with the highest AV in 2016, had a fair salary of $16.13 million
* The most overpaid player (using 'cap hit' as salary metric) between 1994 and 2016 was Tony Romo in 2016, whose fair salary was 0, but whose cap hit was over 20 million
* The most underpaid player (using 'cap hit' again as salary metric) between 1994 and 2016 was Dak Prescott, whose fair salary was 12.3 million, and whose cap hit was 546,000.

## Draft Analyses

The NFL draft is considered the key pipeline for cheap talent because the contract each draft pick receives is fixed in both length and salary (length is four years, salary depends on how high the selection was), due to stipulations in the NFL's CBA.

This is not the case for NFL free agency, where players are free to sign with any team in what resembles a free market.

If we can calculate the fair salary for each draft pick, based on historical AV data for each draft pick, we can then determine how underpaid each draft pick is (which is good for maximizing return on salary cap), if at all.

We begin by creating CSV files for our analysis in R using SQL:

sqlite3 -header -csv nfl\_data.db 'SELECT player.id, player.name, drat.year, draft.round, draft.pick, av.year, av.av\_value FROM av JOIN (player LEFT JOIN draft ON player.id == draft.player\_id) p ON player.id == av.player-id ORDER BY player.id, av.year;' > ../data/draft\_regression.csv

## Setup

Importing the data:

setwd("C:/Users/Steven/OneDrive/Documents/Brown/Junior Year/Second Semester/CSCI1951a/cs1951a\_project/data")  
dr <- read.csv("draft\_regression.csv")  
proj\_salary <- read.csv("rookie\_contracts.csv")  
#dr stands for draft regression

## Data Cleaning

We clean the data, removing undrafted players, players drafted before 1994, and we fix the order of the dataset to have it sorted by draft year, draft pick, and then av\_year, in that order.

Below is the first ten rows of our cleaned table:

## id name draft\_year round pick year av  
## 16829 19941 Dan Wilkinson 1994 1 1 1994 6  
## 16830 19941 Dan Wilkinson 1994 1 1 1995 6  
## 16831 19941 Dan Wilkinson 1994 1 1 1996 7  
## 16832 19941 Dan Wilkinson 1994 1 1 1997 5  
## 16833 19941 Dan Wilkinson 1994 1 1 1998 7  
## 17324 19942 Marshall Faulk 1994 1 2 1994 16  
## 17325 19942 Marshall Faulk 1994 1 2 1995 13  
## 17326 19942 Marshall Faulk 1994 1 2 1996 9  
## 17327 19942 Marshall Faulk 1994 1 2 1997 12  
## 17328 19942 Marshall Faulk 1994 1 2 1998 18

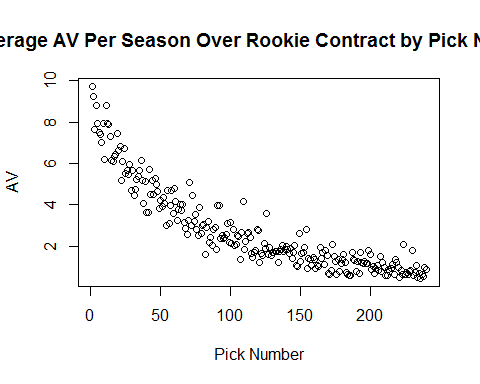
## Aggregating Data into Picks Table

Our data contains draft and av data for players from 1994 to 2016. We want our final product to be a table of all 240 picks, with their actual salary, fair salary, and 'value above fair salary.' 240 is the minimum number of picks in a draft between 1994 and 2016; this allows us to assume there are 23 (or are supposed to be 23) players for every pick, which makes calculations easier. Here we began constructing our Picks Table:

## pick total\_av average\_av log\_av  
## 1 1 894 9.717391 2.273917  
## 2 2 850 9.239130 2.223448  
## 3 3 704 7.652174 2.034990  
## 4 4 809 8.793478 2.174010  
## 5 5 730 7.934783 2.071256  
## 6 6 688 7.478261 2.012000  
## 7 7 682 7.413043 2.003241  
## 8 8 646 7.021739 1.949011  
## 9 9 729 7.923913 2.069885  
## 10 10 570 6.195652 1.823848  
## 11 11 809 8.793478 2.174010  
## 12 12 731 7.945652 2.072625  
## 13 13 724 7.869565 2.063003  
## 14 14 673 7.315217 1.989957  
## 15 15 563 6.119565 1.811491  
## 16 16 561 6.097826 1.807932  
## 17 17 584 6.347826 1.848112  
## 18 18 592 6.434783 1.861718  
## 19 19 683 7.423913 2.004706  
## 20 20 611 6.641304 1.893308  
## 21 21 628 6.826087 1.920752  
## 22 22 476 5.173913 1.643629  
## 23 23 560 6.086957 1.806148  
## 24 24 619 6.728261 1.906317  
## 25 25 508 5.521739 1.708693  
## 26 26 520 5.652174 1.732040  
## 27 27 503 5.467391 1.698802  
## 28 28 547 5.945652 1.782660  
## 29 29 430 4.673913 1.541997  
## 30 30 522 5.673913 1.735879  
## 31 31 412 4.478261 1.499235  
## 32 32 435 4.728261 1.553557  
## 33 33 482 5.239130 1.656156  
## 34 34 492 5.347826 1.676690  
## 35 35 520 5.652174 1.732040  
## 36 36 566 6.152174 1.816806  
## 37 37 475 5.163043 1.641526  
## 38 38 376 4.086957 1.407801  
## 39 39 470 5.108696 1.630944  
## 40 40 335 3.641304 1.292342  
## 41 41 334 3.630435 1.289352  
## 42 42 525 5.706522 1.741610  
## 43 43 413 4.489130 1.501659  
## 44 44 477 5.184783 1.645728  
## 45 45 414 4.500000 1.504077  
## 46 46 487 5.293478 1.666476  
## 47 47 459 4.989130 1.607262  
## 48 48 426 4.630435 1.532651  
## 49 49 352 3.826087 1.341843  
## 50 50 386 4.195652 1.434049

## Regressing on our Data

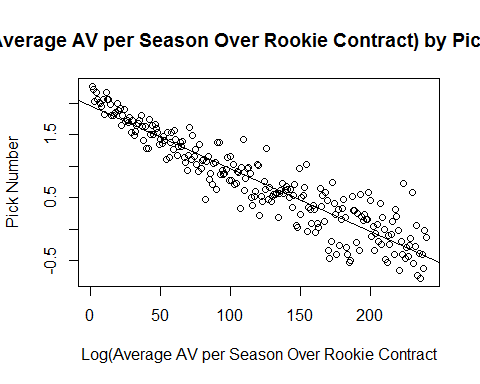
First, let's take a look at our data:



We regress our data using an exponential regression to best fit the data. We get the following regression formula:

We have an R^2 of 0.8676.

##   
## Call:  
## lm(formula = log\_av ~ pick, data = av\_counts)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.71364 -0.16964 0.01007 0.16504 1.00634   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.9563493 0.0350705 55.78 <2e-16 \*\*\*  
## pick -0.0099652 0.0002523 -39.50 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2708 on 238 degrees of freedom  
## Multiple R-squared: 0.8676, Adjusted R-squared: 0.8671   
## F-statistic: 1560 on 1 and 238 DF, p-value: < 2.2e-16



## Fair Salary

Here we calculate the expected AV according to our exponential regression. Using this expected AV, we then calculate the fair salary, assuming that:

1. The salary cap is that of 2017: $167,000,000
2. The salary cap stays the same over the next four years
3. The total AV in all of the next four years is approximated by the average total AV over the last five years (2012-2016), which comes out to 6559.

Assumptions #2 and #3 will almost certainly not be true but we are fairly confident any differences will be marginal; the exact differences are also difficult to predict, so we are fine making these simplifying assumptions.

Below are the fair salaries per year of the rookie contract for the top 25 picks in the draft.

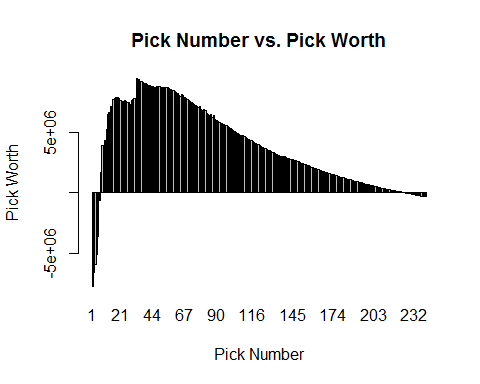
av\_counts$expected\_av = 7.073457\*exp(1)^{-0.0099652\*av\_counts$pick}  
  
pick\_chart = av\_counts[, c("pick", "expected\_av")]  
salary\_cap = 167000000  
total\_av = 6559.4  
num\_teams = 32  
pick\_chart$fair\_salary = pick\_chart$expected\_av/total\_av\*salary\_cap\*num\_teams  
  
head(pick\_chart, 25)

## pick expected\_av fair\_salary  
## 1 1 7.003319 5705664  
## 2 2 6.933876 5649089  
## 3 3 6.865121 5593074  
## 4 4 6.797049 5537615  
## 5 5 6.729651 5482705  
## 6 6 6.662922 5428340  
## 7 7 6.596854 5374514  
## 8 8 6.531442 5321222  
## 9 9 6.466678 5268459  
## 10 10 6.402556 5216218  
## 11 11 6.339070 5164496  
## 12 12 6.276214 5113286  
## 13 13 6.213981 5062584  
## 14 14 6.152365 5012385  
## 15 15 6.091360 4962684  
## 16 16 6.030960 4913475  
## 17 17 5.971158 4864754  
## 18 18 5.911950 4816517  
## 19 19 5.853329 4768758  
## 20 20 5.795289 4721472  
## 21 21 5.737824 4674655  
## 22 22 5.680930 4628303  
## 23 23 5.624599 4582410  
## 24 24 5.568827 4536972  
## 25 25 5.513609 4491985

## Draft Pick Worth

Finally, we have the fair salary for each draft pick per year! We also know that NFL draft picks have relatively fixed salaries. Now, we can calculate the difference between the fair salary of a draft pick, and how much the draft pick is actually paid, to determine the draft pick's theoretical worth ('return above fair salary').

## pick expected\_av fair\_salary estimate total\_fair\_salary pick\_worth  
## 1 1 7.003319 5705664 30566250 22822657 -7743592.6  
## 2 2 6.933876 5649089 29178790 22596355 -6582435.5  
## 3 3 6.865121 5593074 28295878 22372296 -5923582.4  
## 4 4 6.797049 5537615 27286806 22150458 -5136347.6  
## 5 5 6.729651 5482705 25520944 21930821 -3590123.2  
## 6 6 6.662922 5428340 22367629 21713361 -654267.9  
## 7 7 6.596854 5374514 19844976 21498058 1653081.7  
## 8 8 6.531442 5321222 17322323 21284889 3962566.1  
## 9 9 6.466678 5268459 17196047 21073834 3877787.3  
## 10 10 6.402556 5216218 16502457 20864872 4362415.2



## Discussion and Results

Our results confirm the idea that draft picks do tend to be underpaid, and thus are great for maximizing returns on the salary cap. In addition, we find that high first round picks tend to be overpaid, and that the most underpaid pick is actually pick number 33, the top pick of the second round.

These results imply myriad insights (some not necessarily obvious), including: \* The 33rd overall pick is worth more than the 1st overall pick. Yes, you get a worse talent, but you get more bang for the buck. As a General Manager, this also means that trading down in the draft is generally a great strategy. \* In trading for Brock Osweiler this past Spring, the Brown essentially traded 18 million dollars for a second round pick. A mid-second round pick is expected to only provide 8 million dollars in contributions above fair value, implying that this trade was a poor one for the Browns if they valued Brock Osweiler at 10 million dollars or less. In other words, they would have been better off spending that money in free agency.

These are impactful and meaningful insights, and provide substantial promise for future work. However, we note the following limitations (which we could also address in future work):

* The fifth year option for first round picks is not factored into the valuation. For first round picks, teams get a 'option' to add an additional fifth year to the player's contract, usually at a more expensive salary than the previous four years. This past year, 23 of the 32 first round pick options were picked up. Thus, we can expect first round picks to be worth slightly more than what our model suggests. Quantifying this would be a worthy next step.
* Another flaw in our model is that we did not account for possible increases in salary cap. The salary cap has increased substantially over the past 23 years, and it is reasonable to expect it to continue to increase. Accounting for this would increase the 'fair salary' for the draft picks, and further increase the 'worth' of each draft pick. This means that our current estimates for pick worth are almost certainly conservative.
* Finally, it is important to note that Approximate Value is just as its name indicates, an 'approximate' value of a player's contribution. Creating a better metric could be a sizeable undertaking but would potentially improve results.

A final caveat to note is that these draft pick valuations are only valid for future, unknown draft picks. DUring the actual draft, these picks should be instead valued based the players available at each selection (which, of course, is not known until usually a few picks before).