

A Model of Determining Who Gets A Second QB Contract and How Much They Would Get

2024-04-30

Introduction

If we take a bird's eye view of the NFL and zoom out beyond just the players, teams, and games, there always seems to be a lot of discussion surrounding other aspects of football. These other aspects include fantasy football, sports betting, arguing which players are better and who the Greatest Of All Time (GOAT) is, and even the value of players' contracts: if players are getting overpaid or underpaid. When news breaks of massive quarterback contracts, or certain quarterbacks not getting a new contract, fans are always interested in the noise and discussion surrounding NFL contracts and pay.

Therefore, we wanted to tackle the topic of NFL contracts: specifically, finding the likelihood of rookie quarterbacks obtaining a 2nd contract (after their rookie deal), and finding a fair expected value for these quarterbacks' 2nd contracts in today's dollars. This would allow us to be able to predict the odds of future rookie QBs getting a 2nd contract, and be able to see if QBs were overpaid/underpaid by comparing their actual 2nd contract to our modeled "fair" contract (based on performance).

We found this topic to be interesting due to all of the noise and discussion surrounding NFL contracts and pay. Loyal fans of certain star players will always argue that their favorite player is properly paid for all of their work and contribution to a team, while others will argue that those same players are overpaid and not worth that dollar amount. However, fans can debate this back and forth, but a simple way to measure an NFL athlete's contribution, (expected) performance, and overall worth to a team can be boiled down to one number: their contract salary. We thought it would be interesting to take a deeper look at this, and find a fair contract value to help realize if some QBs are overpaid or underpaid.

Data

The contracts dataset we used throughout our analysis is from the `nflreadr` package, and we accessed the dataset using the `load_contracts()` function. This original dataset had 24 columns and ~41,000 rows, each row representing an NFL player with columns including position, team, year signed, contract value, guaranteed contract value, APY, etc.

We started our data filtering/cleaning process by filtering the original dataset for contracts from QBs who were drafted after 1999 (for ease of data), and this narrowed down the selection of players to 240 QBs. We then split this into rookie and non-rookie contracts, and use the non-rookie contracts to find QB 2nd contracts. To look at QB performance stats, we used another dataset from the same `nflreadr` package that we accessed through the `load_player_stats()` function. We then filtered this stats dataset for QBs and grouped the stats by season and QB, calculating/aggregating stats such as completions, attempts, passing yards, passing touchdowns, interceptions, sacks, etc. From then on, we utilized both of our datasets to extract information from rookie QBs who received a 2nd contract and look at their performance stats (completion and interception percentage) from the years before they received their 2nd contract. This gave us a good baseline for rookie QB performance moving forward into our analysis.

Below are some data visualization graphs of some exploratory data analysis we did throughout our study, including a look at rookie player stats, passing touchdowns vs. interceptions (showing 81% of rookie QBs got a 2nd contract), and inflated annual contract value.

```
## # A tibble: 41,007 x 24
##   player      position team is_active year_signed years value  apy guaranteed
##   <chr>      <chr>  <chr> <lgl>      <int> <int> <dbl> <dbl>      <dbl>
## 1 Joe Burrow  QB      Beng~ TRUE      2023     5 275  55      219.
## 2 Aaron Rodg~ QB      GB/N~ FALSE    2022     5 151.  50.3    151.
## 3 Josh Allen  QB      Bills TRUE      2021     6 258  43      150
## 4 Russell Wi~ QB      Bron~ FALSE    2022     5 245  49      165
## 5 Justin Her~ QB      Char~ TRUE      2023     5 262.  52.5    194.
## 6 Lamar Jack~ QB      Rave~ TRUE      2023     5 260  52      185
## 7 Patrick Ma~ QB      Chie~ TRUE      2020    10 450  45      141
## 8 Jalen Hurts QB      Eagl~ TRUE      2023     5 255  51      179.
## 9 Kyler Murr~ QB      Card~ TRUE      2022     5 230.  46.1    160
## 10 Deshaun Wa~ QB      Brow~ TRUE      2022     5 230  46      230
## # i 40,997 more rows
## # i 15 more variables: apy_cap_pct <dbl>, inflated_value <dbl>,
## #   inflated_apy <dbl>, inflated_guaranteed <dbl>, player_page <chr>,
## #   otc_id <int>, date_of_birth <chr>, height <chr>, weight <chr>,
## #   college <chr>, draft_year <int>, draft_round <int>, draft_overall <int>,
## #   draft_team <chr>, cols <list>
```

Slice function source: <https://dplyr.tidyverse.org/reference/slice.html>

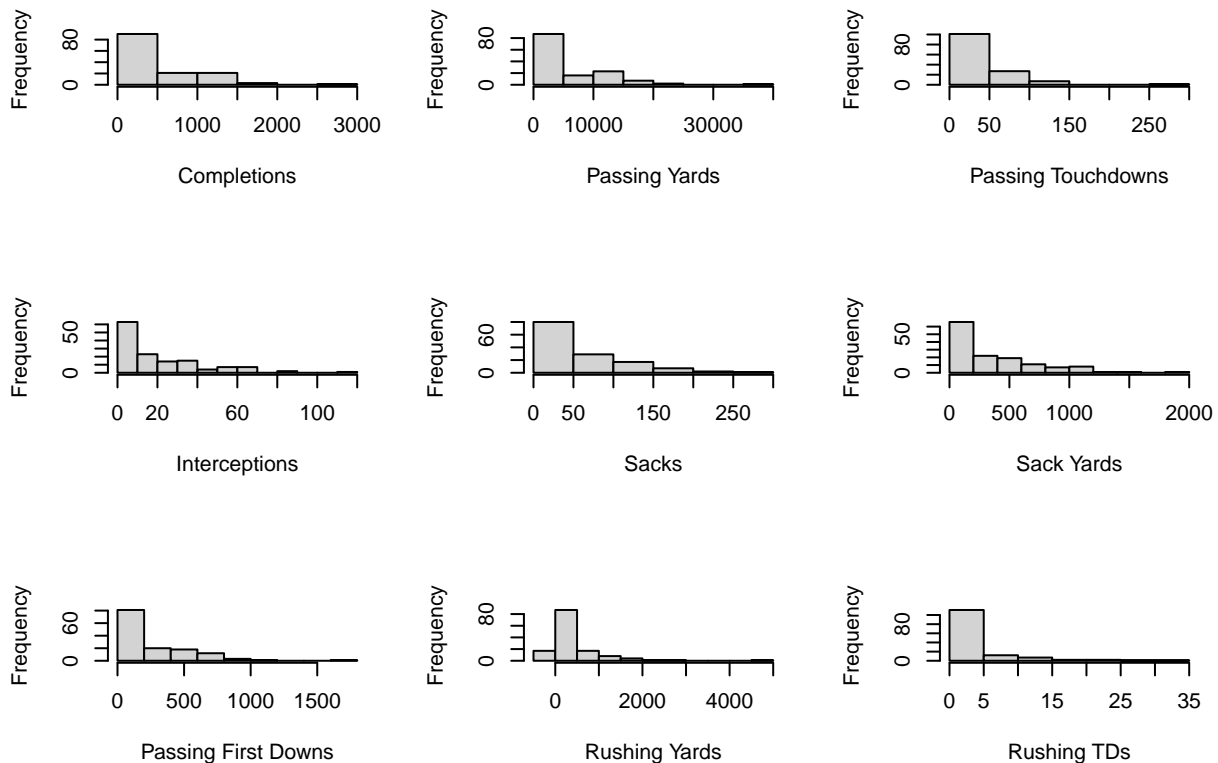


Figure 1: Plot of Distributions of relevant rookie contract metrics

Passing Touchdowns vs. Interceptions Scatterplot

Red Indicates QB received 2nd contract, Blue indicates otherwise

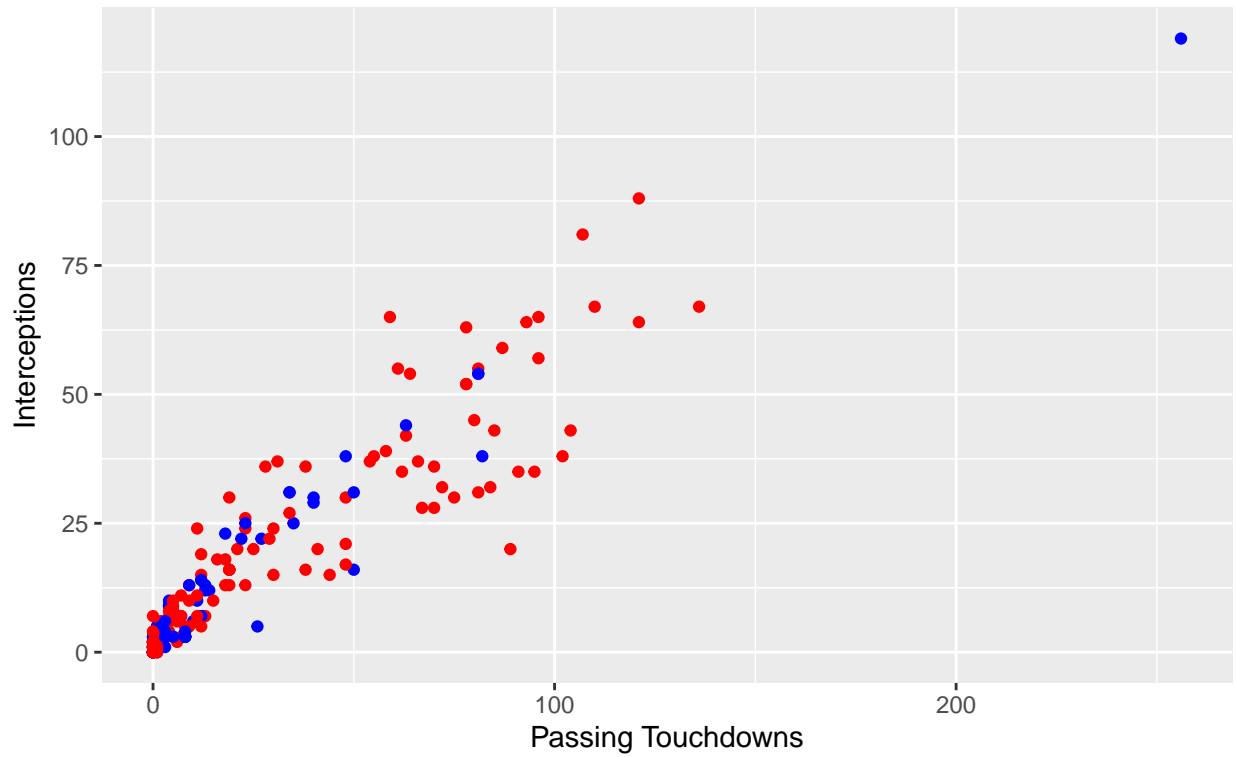


Figure 2: Plot of Passing Touchdowns vs Interceptions and Whether A QB Got a Second Contract

Histogram of Second QB Contract Annual Value
Values Are in 2024 USD

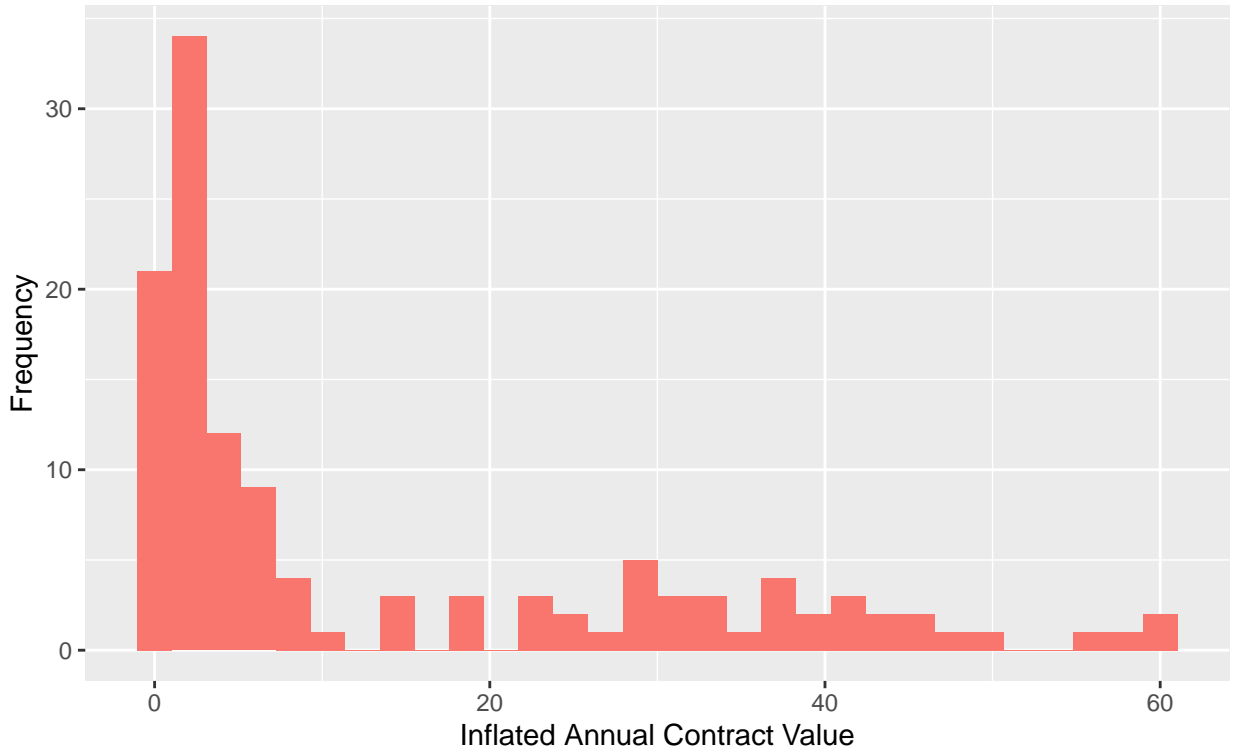


Figure 3: Inflated Annual Contract Values Histogram of Second Contracts

Methods

Model Building

To reiterate the two questions we seek to answer:

- 1) What is the probability that a rookie quarterback receives a second contract given their performance on their first contract?
- 2) What is the fair value of the second contract a rookie QB receives, given their performance on their first contract?

Due to the different natures of the two research questions, we will build two different models, one for each question.

In the preliminary stages, we decided to build a bunch of models for each question, then perform k-fold cross validation to determine the best one. As such, we built 4 models for the first question and 4 different models for the second question. For the first question we built a binomial GLM with all relevant variables, a GAM with all relevant variables, a binomial GLM without rushing metrics and a GAM without rushing metrics. For the second question, we built a GLM with all relevant variables, a Kernel model with all relevant variables, a GLM without rushing metrics, and a Kernel model without rushing metrics.

We chose a binomial GLM model for our first question because we are trying to predict the probability of a rookie receiving a second contract given their performance. Since this question is probabilistic in nature, a binomial GLM model would be a good, simple model to predict a binary outcome, either a QB receives a

second contract or they don't. The GAM model has the same overall goal as the binomial GLMs, but we decided to use a GAM to account for potential non-linear patterns and effects of predictor variables on the outcome variable.

The binomial GLM assumptions we can reasonably assume to be met are that the rookie QBs and their performance are independent of each other, the outcome variable is binary (either the QB receives a second contract or they don't). The GAM assumptions that we can reasonably assume to be met are that the variables are additive, so the effect of each predictor on the response variable is independent of the other variables, the relationship between each predictor and the response is linear, and the rookie QBs are independent of each other.

For our second question, we chose a GLM because we are now trying to predict contract amount, which is a dollar amount. We wanted to start with a basic model that is close to a conventional linear regression model, before building more complex models. We chose a Kernel regression model for the flexibility it provides, as it can capture complex, non-linear relationships between variables.

As for the assumptions, we can reasonably assume that the rookie QBs are independent of each other and that there is a roughly linear relationship between the response and predictor variables. The assumptions we have for the Kernel model are that the rookie QBs are independent of each other.

Model Evaluation

Now that we've built our models, we want to evaluate them and see which one performs the best for each question. The general evaluation technique we used is k-fold cross validation with 4 folds. For the binomial GLM and GAM, we used k-fold cross validation to calculate mean Brier Scores \pm 1 SD for each model. We chose to use Brier Scores as a way to evaluate model performance for the first question because we are predicting the probability that a GB receives a second contract. For the GLM and Kernel, we used k-fold cross validation to calculate MSE \pm 1 SD for each model, since we are trying to predict the amount a QB is worth. In general, we wanted to pick the model with the lowest mean Brier Score and the lowest MSE, with the values being calculated through 4-fold cross validation.

After evaluating our models, we determined that for the first research question, the GAM model without rushing metrics performed the best with the lowest Brier Score of 0.124. For the second research question, the Kernel model without rushing metrics performed the best with the lowest MSE of 36.29.

Uncertainty in Model Estimates

We turned to bootstrapping to quantify the uncertainty in the models we ultimately decided upon. We know the GAM model without rushing yards performed the best for the first question, and the Kernel model without rushing yards. We bootstrapped these two models to calculate confidence intervals that gives us a low and high bound for the uncertainty around the probability of a rookie QB receiving a second contract and the fair value of their second contract.

```
##
## 1 2 3 4
## 34 34 34 34
```

```
##
## 1 2 3 4
## 34 34 34 34
```

```
##
## 1 2 3 4
## 31 31 31 31
```

```
##
## 1 2 3 4
## 31 31 31 31
```

```
-----
x
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 0.0957 0.135 0.174
-----
```

Figure 4: Brier Score for Kitchen Sink Model Under 4 fold CV +/- 1 SE

```
-----
x
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 0.0868 0.124 0.161
-----
```

Figure 5: MSE Scores for Kitchen Sink Model to Predict Contract Value Under 4 Fold CV

```
-----
x
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 0.108 0.142 0.177
-----
```

Figure 6: Brier Score +/- 1 SE Under 4 fold CV for Passing Model

```
-----
x
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 0.0916 0.128 0.164
-----
```

Figure 6: Brier Score for GAM Model and Passing Only +/- 1 SE

```
-----
x
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 74.1 86.4 98.7
-----
```

Figure 7: Brier Score for GLM Model Kitchen Sink +/- 1 SE

```

-----
x
-----
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 135. 309. 483.
-----

```

Figure 6: CV MSE for Kernel Model +/- 1 SE

```

|x
|:-----|
|# A tibble: 1 x 3
|'Lower Score' 'Median Score' 'Upper Score'
|<dbl> <dbl> <dbl>
|1 84.3 97.4 111.
|

```

****Figure 6: Brier Score for GAM Model and Passing Only +/- 1 SE****

```

-----
x
-----
# A tibble: 1 x 3
  Lower Score Median Score Upper Score
1 7.22 22.6 37.9
-----

```

Figure 6: Brier Score for GAM Model and Passing Only +/- 1 SE

Used code from Homework 3 solutions

Results

Binomial Model Interpretations

In comparing our binomial models' performances, regarding odds of receiving a second contract, we chose to cross validate and calculate the Brier Scores across a 4 fold process. We could then ascertain the model with the strongest predictive ability based on their resulting scores, as a lower Brier Score suggests that our predicted probability was not far off from the measured probability. To this end, we found the GAM model comprising the passing metrics (passing_yds, passing_tds, sacks, passing_first_downs) to have the best Brier score of 0.127. To give an idea for the range of our outcomes, the highest Brier score we encountered was in that of our GLM kitchen sink model at 0.142. On a possible score range of [0,1], we are more than pleased with our model's ability to predict whether a rookie quarterback will receive a second contract. Furthermore, when examining the partial response graphs for the kitchen sink model, the predictors regarding the passing metrics had the least spread among the rest, suggesting they would be reliable. Having said, the passing touchdowns and passing first downs did have a positive correlation to increasing the odds of a contract. However, we saw an inverse relationship and then a positive relationship from passing yards and sacks taken, respectively, which had the opposite effect on the second contract odds than what we would have expected.

Contract Value Model Interpretation

Then when it came time to compare our contract models' performances, regarding expected payoff of a theoretical second contract, we chose to cross validate and calculate the MSE across a 4 fold process. We could then ascertain the model with the strongest predictive ability based on their resulting values, as a lower MSE suggests that our predicted values on average were closer to the real values. To this end, we found the npreg Kernel model using the predictors of only completions, passing_yards, passing_tds, interceptions, sacks, sack_fumbles_lost, and passing_first_downs to have the best MSE value of 36.27. To give an idea for the range of our outcomes, the highest MSE during one cross validation was 330.27 we encountered was in that of our kernel estimator kitchen sink model. With the ever possible risk of overfitting a model even with our train and testing data, it is reassuring to see some of the noise predictors be filtered out through our model valuation process.

Possible Uncertainty

With our methodical approach to data cleaning, cross validating, training + testing, and model selection, we feel pleased with the predictive capabilities of our models given the noise of real world data and the limited number of quarterback contracts we could be working with. To test these out, we bootstrapped 95% confidence intervals for both the probability of 4 current quarterbacks on rookie deals to receive another contract with the best performing binomial model, as well as bootstrapping a 95% confidence interval for a player's expected APY in a hypothetical second contract. The results were consistent with what we would expect, with higher performing quarterbacks like Trevor Lawrence getting higher probabilities and expected values than lower rated quarterbacks like Kenny Kickett and Mac Jones.

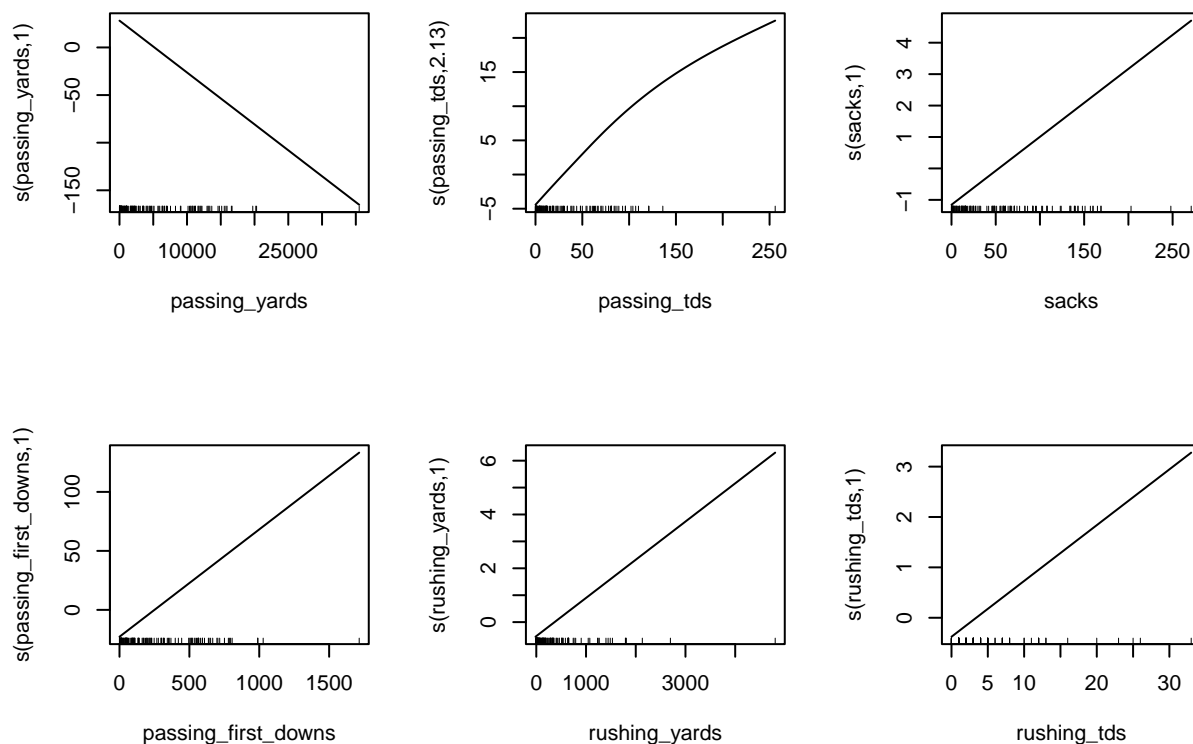


Figure 10: Partial Response Function for GAM

```

x
# A tibble: 4 x 2
Odds Second Contract Expected APY(M)
* <dbl[1d]>
1 0.541 7.56
2 0.243 19.8
3 0.891 27.8
4 0.972 36.9

```

	Lower Bound	Upper Bound
Kenny Pickett	-7240.218	2.0812
Mac Jones	-3801.5755	4740.0201
Trevor Lawrence	-3809.5045	16024.3495
Tua	-3477.1184	8919.7084

Lower Bound and Upper Bound for Odds of Second Contract via Bootstrapping

	Lower Bound	Upper Bound
Kenny Pickett	-23.0969	13.2805
Mac Jones	-9.2836	37.5087
Trevor Lawrence	10.8992	54.615
Tua	14.9489	70.1135

Lower Bound and Upper Bound for Predicted Value of Second Contract via Bootstrapping
Lower Bound and Upper Bound for Expected APY via Bootstrapping

Predicted Odds of Second Contract vs. Annual Contract Value
Quarterbacks Drafted 2020 or later

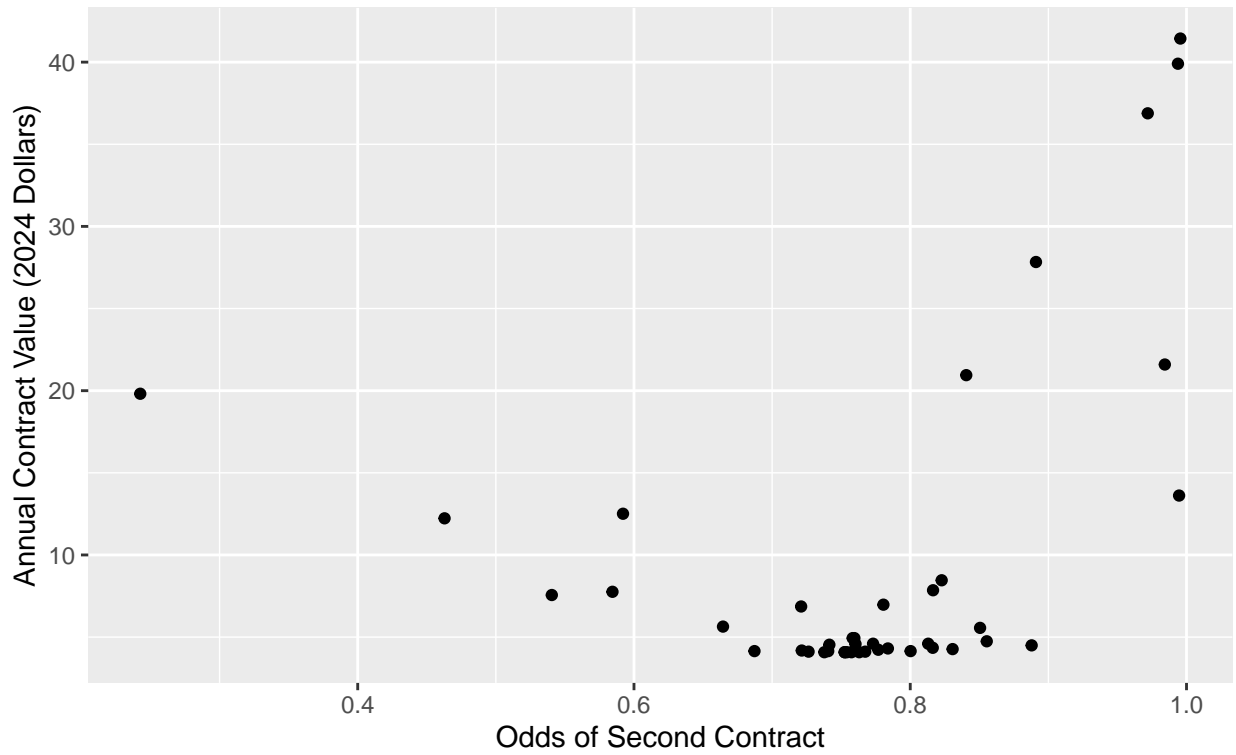


Figure X: Predicted Odds of Second Contract Along with Expected APY for QBs Drafted 2020 or later

Bootstrapping code copied from 36-402 chapter 6 of Cosma Shalizi's textbook

##Discussion

Our two optimized models with decently strong predictive abilities suggest that predictors related to a player's passing are the most likely to translate into them receiving a second contract, while the expected APY kernel model of such a contract takes into account more holistic stats from the player. Given the multidimensionality of the data in the kernel and possible correlation among predictors, there could be relationships we have not accounted for in this first sports data analysis, and as trends in quarterback performance change with the league, we would expect our adaptive training method to be able to stay relevant.

There are some limitations that we want the reader to be aware of. Namely, the data is not all encompassing, we only looked at data after 2020, which is a narrow stretch in the grand scheme of NFL contracts. Additionally, our data only looks at quantifiable performance metrics, not qualitative details like player injury history or personality. We acknowledge that player chemistry on the field can impact how players perform, and past injuries can rear up and suddenly cause a drop in performance. Team dynamics overall also play a large role in how QBs perform, and we were unable to capture that in our model.

In the future, we hope to take our methodology and apply it to other contracts beyond the second contract, as well as other positions beyond the QB position. It would also be interesting to quantify team dynamics, although this appears difficult to do, and requires more thought on how to do this without losing valuable qualitative data.