UiT The Arctic University of Norway

School of Business and Economics

Seeking causal effects from a rule change

How did changing the “roughing the passer rule” in 2018 affect salaries for quarterbacks in the NFL?

Thomas Tolla Jakobsen

Master’s Thesis in Economics, SOK-3901-01, June 2021

Acknowledgments

Finishing this thesis concludes my education at the School of Business and Economics at UIT The Artic University of Norway. My experiences have been of great value, both educationally and socially.

I would like to start with thanking my supervisor, Øivind Devik Schøyen. First and foremost for your ability to guide me through the process by letting me try and fail, as it would give me the most valuable learning experience. In addition to this, our endless and insightful discussions have been invaluable to me.

During my two last years here in Tromsø, my classmates have been of immense importance to me. Our coffee breaks, table-tennis sessions and all the laughter in our office have made my time here very cheerful.

At last I need to thank my family and friends for their support through all my years at the University. A special thanks goes out to my mother for proofreading most of my assignments through the years.

Thomas Tolla Jakobsen

Abstract

Following concerns for players safety, the National Football League (NFL) chose to protect the quarterbacks (QB) further by changing the “roughing the passer rule” in 2018. This paper studies the effect this rule change had on QB salaries. This is done by applying a Difference-in-Differences approach on salary data stretching from 2013 to 2021. Our estimate show how the salary increase following the rule change around 900 thousand dollars when excluding bonuses, while the actual increase reach 1.2 million dollars. Considering the wages for the top 30 percent of QBs, we find an increase of actual spending approximately reaching 3 million dollars, while for the top five percent the increase of actual spending lies at around 1.3 million dollars. Controlling for time trends gave us answer similar to our original DID approach, but our robustness checks gave us treatment effects similar to the ones from our DID approach. We conclude that since there are treatment effects, even when there are no treatment, we cannot accurately state that it is the rule change that creates the treatment effect.

Keywords: NFL, QB Salaries, Pareto Distribution, Differences-in-Differences

Table of Contents

List of Figures:

List of Tables:

1. Introduction

In 2018 The National Football League (NFL) introduced a change in the "*roughing the passer rule"*. The "rouching the passer rule" is referred to as RULE 12 SECTION 2 ARTICLE 11 in the NFL rulebook and will be referred to as RTP (Operations, 2018). The original rule (from 1995), stated that: "a defensive player is prohibited from unnecessarily and violently throwing him down **and** landing on top of him with all or most of the defenders weight." The update (implemented in the 2018 season), stated that: "a defensive player is prohibited from unnecessarily and violently throwing him down **or** landing on top of him with all or most of the defenders weight." The defenders job is to prevent the quarterback (QB) from throwing the ball and this is done by "sacking" the QB, so changing the rules makes it harder for the defender to "sack" the QB.

The change from **and** to **or** is a small one, but the consequences were huge. In 2020 there was an average of 3.97 RTP-penalties called for each of the 32 teams, while in 2019 there was an average of 4.25 RTP-penalties. However, back in 2016 the average were down at 2.78 RTP-penalties (*NFL Penalties*, 2020). The increased frequency, at which those penalties are enforced, is a good starting point.

Our argument therefore becomes; since the penalties enforced as a result of this rule change has increased, would this imply that the job of the defenders becomes harder? And in the same sense, that the job of the QB has become easier? The aim of this paper is to draw a line between the increased success of the QB as a result of the changed RTP-rule, and the following increased salary of the QB.

Furthermore, common knowledge assumes that QBs affect the game to a higher degree than other positions (Schalter, 2013). This assumption leads to the conclusion that the upper level of players in the NFL are QBs, while the rest of the players are at the level below. This paper will try to expand on this knowledge, and furthermore say that the salary of players in the NFL follow a certain Pareto distribution. The term is named after the famous economist Wilfredo Pareto, who insisted that in societal structures, workers with different set of skills earn different salaries, depending on how highly their skills were valued (Pareto, 1896). Going a step further, we will try to show the existence of this Pareto distribution in the NFL, while at the same time trying to show that the rule change further increased the gap between QBs and all the other positions.

2. Theory

Researchers from the University of Hamburg had a useful summary comprising of the most important aspects of the contract negotiation between a team and a player (Heubeck, 2003). The salary for a given player consists of three parts;

(1) *signing bonuses* – these are payments set in place for a player when signing a new contract. If the team chooses to terminate the contract, this payment is still guaranteed to the player.

(2) *fixed payments* – this part of the salary is a yearly payment given to the player, and usually count against the teams Salary Cap. The Salary Cap is a cap that is set by the league. This cap determines how much each team is allowed to spend on salaries, and is meant to create a level playing field between different teams.

(3) *performance bonuses* – incentives based on either personal performance or team performance.

One important note to labor economics is the existence of collective bargaining between owners and unions. With the latter seeking an acceptable wage for its members, while the firms have an incentive to depreciate the value the union sets forward. The most commonly known model for evaluating collective bargaining, is referred to as the “right-to-manage” model (Cahuc, 2014). This is also applicable to the NFL, where the union for players is called the National Football League Players Association. The NFLPA acts on behalf of the players when it comes to bargain over the collective bargaining agreement or CBA. The last CBA was implemented before the 2020 season of the NFL(NFLPA, 2020).

While we are not interested in the contract negotiation, it still seems important to note as it is a prerequisite to the contracts negotiated.

2.1 Previous literature

The NFL has been a research subject for economists in a few years now, and sports economists have chosen this subject as it translates well into other parts of society.

In 2019 two researchers from the University of Pennsylvania research how teams could better allocate the money they had available to them (Mulholland & Jensen, 2019). Mulholland and Jensen found that teams who chose to draft players that would have an immediate impact on the team’s performance, had allocated their money in the best manner. This relates well to our research question, as we have a presumption of QBs high impact on the team’s performance.

Another important research paper stems from Michael A. Roach from Middle Tennessee State University, in which the researcher tests the NFL for labor market efficiency with specific position groups (Roach, 2017). The researcher ends up concluding that the teams choosing to spend more money on the offensive side of the ball, mainly QBs and offensive line (the ones who protect the QB) will earn the highest value for their money.

Both articles provide similarities to our intuition, which is that teams who choose to spend more money on QBs, will in return receive the highest value from their spending.

2.2 Distribution of wealth in the NFL

We center our research behind the assumption that QBs will become more valuable as the rule change is in their favor. But is there already some sort of disparity between athletes in the NFL? If there is, a summary of this Pareto distribution will prove useful.

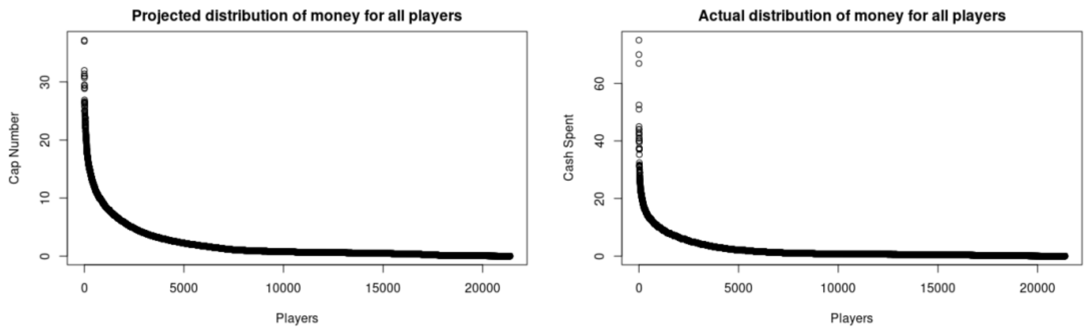

Figure 1: Projected and actual distribution of money for all players.

Notes on Figure 1: Cap Number and Cash Spend show dollars in millions on the y-axis, while the number of players are listed on the x-axis.

Figure 1 shows us how there is a few top earners in the league, while the average amount of players lies below. Cap Number is the amount of money that counts against said year Salary Cap, while Cash Spent reveal what the salary of a certain player actually was. Deciding how much goes against the Salary Cap, is down to the individual team, and given the fact that bonuses heavily influence the actual cash given out to a player, Cash Spent is almost always larger than Cap Number. It goes the other way around too; if a player is injured, Cap Number stays the same, while Cash Spent decreases.

While it is interesting to check for Pareto distribution among all players in the NFL, we think that for our thesis, it would be more interesting to look at QBs alone. We do this by creating three different Pareto distributions, one for all QBs, the next one for the top 30 percent and the last one for the top five percent of QBs. The top 30 percent of Cap Number are for example the 286 highest Cap Numbers out of 954 QB observations for Cap Number. The rest of the figures follows this same notion.

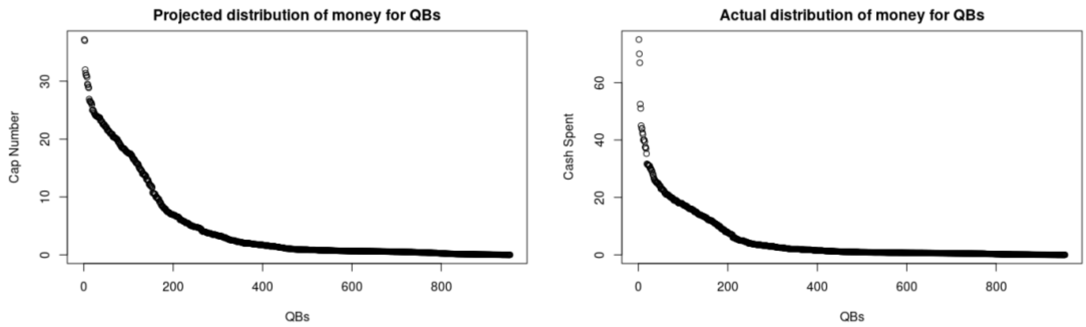

Figure 2: Projected and actual distribution of money for QBs.

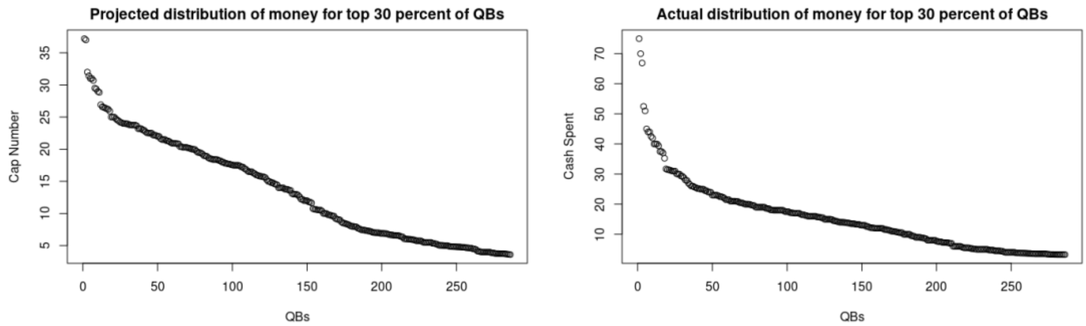

Figure 3: Projected and actual distribution of money for the top 30 percent of QBs.

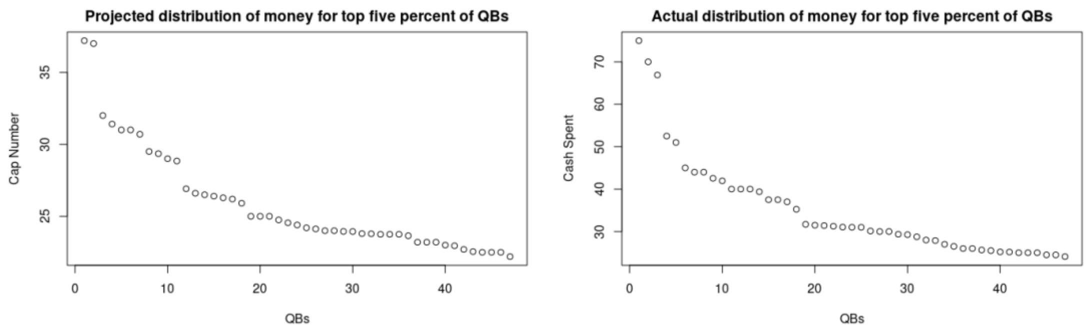

Figure 4: Projected and actual distribution of money for the top five percent of QBs.

Notes on Figure 2, 3 and 4: Cap Number and Cash Spend show dollars in millions on the y-axis, while the number of QBs are listed on the x-axis.

In the figures above, the notion we implied earlier still sticks. The Pareto distribution still exists, and this would further imply that even among the QBs themselves there is some degree of salary dispersion.

The existence of disparity even between the QBs, allows us to not only check for causal effects between the rule change and increasing salary for QBs, but it also enables us to check if the increase in salary applies to the highest earners among all QBs.

Going back to the research question and introducing the hypotheses that will be tested:
How did changing the "rouching the passer rule" in 2018 affect salaries for quarterbacks in the NFL?

- H1: As the rule change is enforced, this will result in an increasing salary for QBs.

3. Method

3.1 Data collection

To investigate if there exits causal effects between the rule change and the following increase in salaries for QBs, information about teams Cap Number and Cash Spent would prove useful. Achieving this was done by the use of web scraping from an open source called "overthecap.com" (*PS*, 2021).

Table 1: Descriptive statistics of the data.

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Max
Cap Number	21,372	1,961,271	3,219,914	37,202,000
Cash Spent	21,372	2,136,586	3,837,446	75,000,000
Salary Cap	21,372	164,498,844	24,100,096	198,200,000

Notes on Table 1: Summary of important numerical variables in our data set. Scraping players specific data stretching from 2013 to 2021, allows us to end up with a large sample of players. The numbers are listed in dollars. Binary variables such as QB and RI were left out of our descriptive statistics.

QB is our first indicator variable; it takes the value one for all observations that are QBs and zero otherwise. The correlation in both instances (Cap Number and Cash Spent) with regards to the binary variable QB, shows a low degree of correlation (0,18 and 0,17). RI is the next dummy variable. This variable will be equal to one for all observations after 2018 and zero before 2018. Additionally, we included a numerical variable for "year".

We conclude that the data is comprised as a panel, however an unbalanced panel. The prerequisite for doing a Difference-in-Differences analysis, does not hinge on the panel data being balanced, but on the fact that there is a clear intervention and a result following that intervention. We therefore feel comfortable continuing our process with this data.

3.2 Difference-in-Differences

Difference-in-Differences, or DID, is a research method that allows us to check for causality between an intervention and what that intervention entails for the group affected (Hill et al., 2018). For our purposes, DID will be used to prove that the rule change is the main driver behind the increasing salary for QBs.

Our experiment assumes that the Treatment group will be QBs, while all the other players will be the Control group.

We choose to create two regression equations; one with Cap Number as the dependent variable, and one with Cash Spent. The causal effects or Treatment effect estimated by DID will be noted as the δ-coefficient, and this coefficient shows the change that occurs when the rule is implemented. This coefficient is the interaction between the independent variables QB and RI. Both equations and the proof of why the δ-coefficient is the one for inference about the change induced from the rule change, can be found in the Appendix (as equation (1) and (2)).

We also want to prove that DID could be estimated using sample means. Using sample means to estimate the δ-coefficient was introduced by David Card and Alan B. Krueger, in their paper concerning minimum wages in both Pennsylvania and New Jersey (Card & Krueger, 2000).

When creating sample means, we first need to add some definitions:

$\boldsymbol{Y}$ = Our outcome variable (either for Cap Number or Cash Spent)

After defining Y as our outcome variable, we will use four different sample means for finding the δ-coefficient:

$\boldsymbol{Y_{TB}}$ = Sample mean of Y for the Treatment group, before intervention.

$\boldsymbol{Y_{TA}}$ = Sample mean of Y for the Treatment group, after intervention.

$\boldsymbol{Y_{CB}}$ = Sample mean of Y for the Control group, before intervention.

$\boldsymbol{Y_{CA}}$ = Sample mean of Y for the Control group, after intervention.

Creating the sample means for both the Treatment- and Control group, allows us to create two equations for finding the δ-coefficient for both Cap Number and Cash Spent:

$$\boldsymbol{\delta_{CapNumber}} = [(Y_{TACN} - Y_{CACN}) - (Y_{TBCN} - Y_{CBCN})] \tag{3}$$

$$\boldsymbol{\delta_{CashSpent}} = [(Y_{TACS} - Y_{CACS}) - (Y_{TBCS} - Y_{CBCS})] \tag{4}$$

In addition to finding the mean values for DID, we also want to prove our results really stem from the rule change in 2018. We do this by implementing a Placebo test that lets us check whether or not a false rule change implemented in either 2017 or 2019, gives significant results. The last thing we do is to check if the salary of the top percent of QBs act in the same way as QBs altogether. We do this by first checking for the top 30 percent of QBs and at the end checking for the top five percent of QBs. Additionally, we introduce a time trend for both the Treatment- and Control groups. This time trend will be further used to control the answers from our original DID-estimation.

4. Results

4.1 Time trend

Before estimating DID, we want to check for a time trend by using the variable "year", on the effects for both Cap Number and Cash Spent. This is done by applying creating sample means of Cap Number and Cash Spent for both the Treatment- and Control groups:

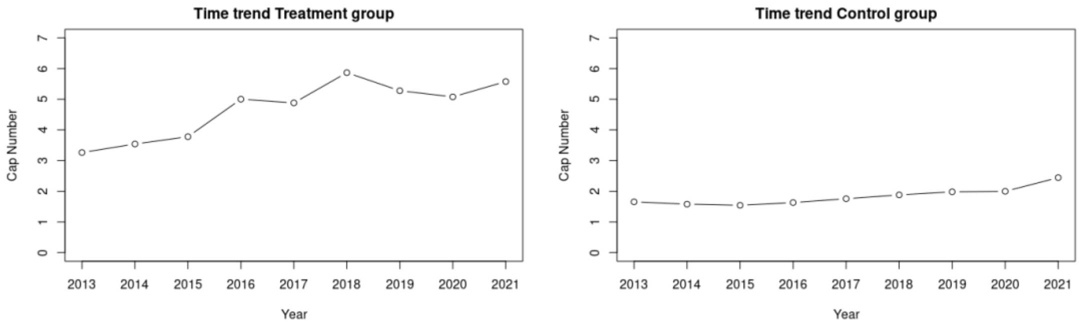

Figure 5: Time trends for the Treatment- and Control group with Cap Number as the dependent variable.

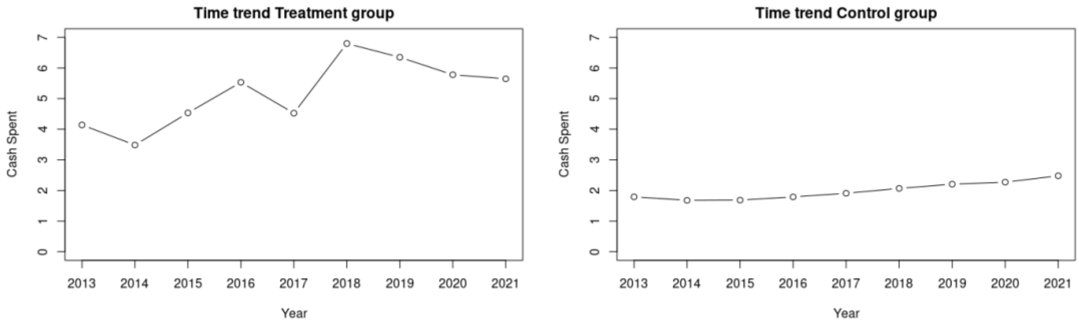

Figure 6: Time trends for Treatment and Control group with Cash Spent as the dependent variable.

Notes on Figure 5 and 6: Cap Number and Cash Spent show dollars in millions on the y-axis, while the years are pictured on the x-axis.

There is clearly a difference in trend for the Treatment- and Control group. The is a higher acceleration in increasing Cap Number and Cash Spent for the Treatment group than the Control group. Additionally; the increase from 2017 to 2018 for the Treatment group, makes us vary of finding significant results for Treatment effects stemming from the 2018 rule change.

Economists tends to use time trend to show that without the intervention, both the Treatment group and the Control group acts in a similar way. Though, this is not a prerequisite, as we already mentioned that there are underlying differences between those who play QB and those who do not. Later in the results section, we will show how controlling for time trends affect our regression estimates for DID.

4.2 Results DID

Table 2: Regression output for DID.

Differences-in-Difference

	Dependent variable:	
	Cap Number	Cash Spent
	(1)	(2)
QB	2,465,662***	2,670,573***
	(140,817)	(168,197)
RI	433,328***	478,992***
	(44,306)	(52,921)
QB*RI	897,399***	1,208,367***
	(210,119)	(250,976)
Constant	1,634,213***	1,773,181***
	(30,043)	(35,885)
Observations	21,372	21,372
Adjusted R^2	0	0
F Statistic (df = 3; 21368)	296***	263***
Note:		*p**p***p<0.01

Notes on Table 2: Standard errors are in parenthesis and “QB*RI” explains the δ-coefficient. The low adjusted R^2 indicates that there is a lot of unexpected variance in our variables. However, we still feel comfortable saying that there is a significant value for change in QB salaries, as a result of the rule change.

The result is; teams plan to spend 897,399 dollars more on QBs after the rule change, but they actually spend 1,208,367 dollars more on QBs as a result of the rule changing. This means that our hypotheses surrounding QBs checks out, and now we will go on to further visualize and control this result.

4.3 Visualization of the results

Often it is easier to understand results, if we are shown visuals of them. Using this notion, we choose to present the Treatment, Control and the Counterfactual Treatment in two figures. The Counterfactual Treatment is an estimation using the trend of the Control group, with the same starting point as the Treatment group. We will use these notations in our visualization:

- TA = Treatment After Intervention
- TB = Treatment Before Intervention
- CA = Control After Intervention
- CB = Control Before Intervention
- CTA = Counterfactual Treatment After Intervention

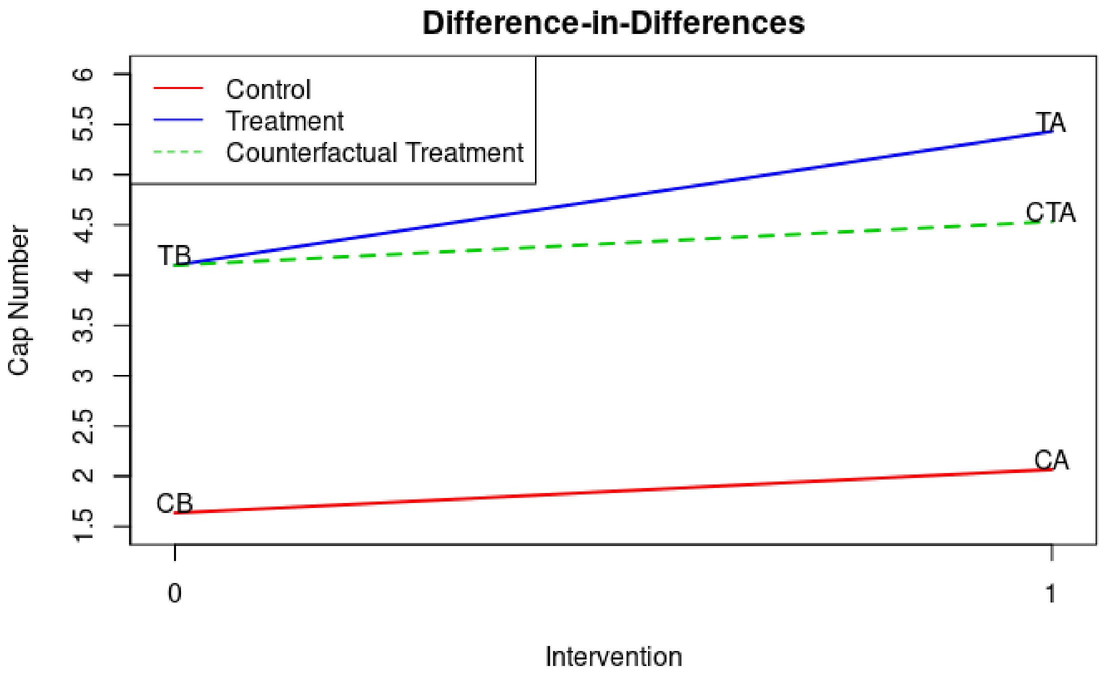

Figure 7: Visualization of DID for Cap Number.

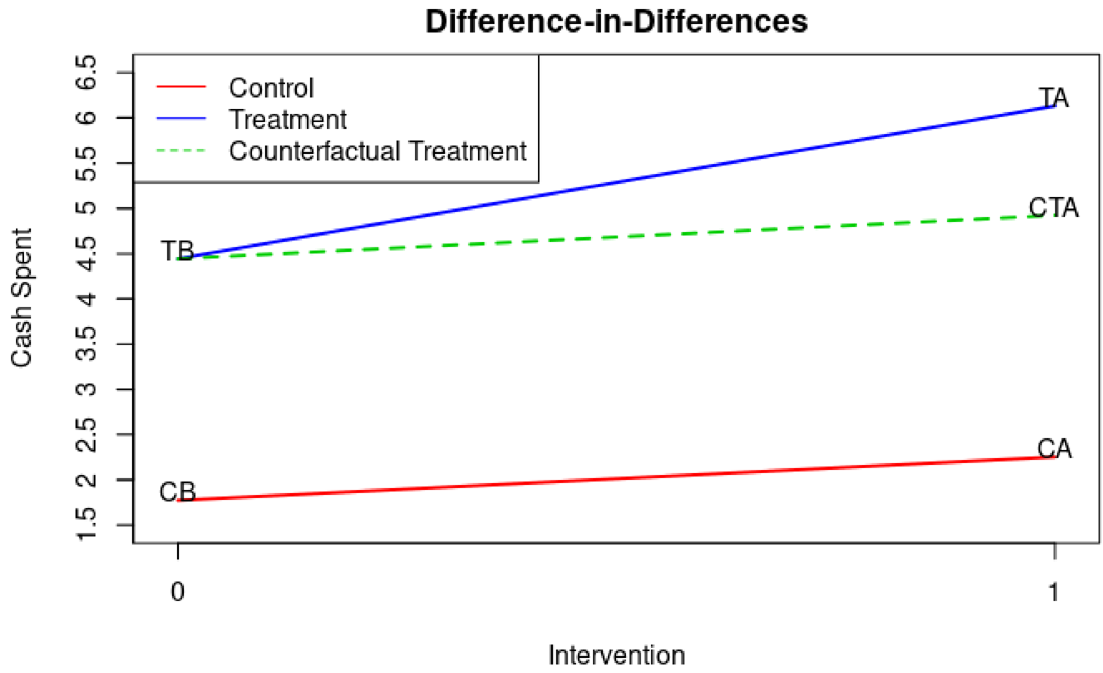

Figure 8: Visualization of DID for Cash Spent.

Notes on figure 7 and 8: Cap Number and Cash Spend show dollars in millions on the y-axis, while intervention emulates the period before and after the rule is implemented, hence 0 is before and 1 is after. In Figure 7 the difference between TA and CTA is the δ-coefficient (897,399), while in Figure 8 the same notion applies (1,208,367).

Showing visuals allows us to see the Treatment effect (δ) better. It shows that the difference between the end of Treatment (TA) and end of Control (CA) is not the difference we are looking for. This is because the underlying differences between the Treatment- and Control group are not accounted for here. Instead, using the Counterfactual Treatment (CTA), we find the true Treatment effect (δ). In other words, using the difference between TA and CTA allows us to eliminate the underlying differences between the Treatment-and Control group.

4.4 Results using sample means

Another way of proving the DID, is using sample means before and after the rule change.

Table 3: Sample means for Cap Number and Cash Spent.

Variable	Mean Cap Number	Mean Cash Spent
Treatment After	5,430,602	6,131,114
Treatment Before	4,099,875	4,443,754
Control After	2,067,541	2,252,173
Control Before	1,634,213	1,773,181
δ	897,399	1,208,367

Notes on Table 3: Sample means for Cap Number and Cash Spent in dollars. In addition to the mean values, we choose to include the Treatment effect (δ), we arrive at if we use equation 3 and 4 from the method part of the article.

Using sample means is a way of showing the same as with our DID-regression:

$$\boldsymbol{\delta_{CapNumber}} = [(Y_{TACN} - Y_{CACN}) - (Y_{TBCN} - Y_{CBCN})]$$

$$\boldsymbol{\delta_{CapNumber}} = [(5{,}430{,}602 - 2{,}067{,}641) - (4{,}099{,}875 - 1{,}634{,}213)] \tag{3}$$

$$\boldsymbol{\delta_{CapNumber}} = 897{,}399$$

$$\boldsymbol{\delta_{CashSpent}} = [(Y_{TACS} - Y_{CACS}) - (Y_{TBCS} - Y_{CBCS})]$$

$$\boldsymbol{\delta_{CashSpent}} = [(6{,}313{,}111 - 2{,}252{,}173) - (4{,}443{,}754 - 1{,}773{,}181)] \tag{4}$$

$$\boldsymbol{\delta_{CashSpent}} = 1{,}298{,}367$$

Notes on equation 1 and 2: We simply inserted the actual sample means from our data. By using the same equations as presented in the method section, we arrive at the same values for the Treatment effect (δ).

4.5 Controlling for time trends, top 30 & five percent of QBs and robustness checks

We previously mentioned the existence of an increasing time trend for Cap Number and Cash Spent. Adding the time trend to our regression, allows us to control our δ-coefficient with regards to said time trend. In addition to this we wanted to check the same coefficient for the top 30 and five percent of QBs, and at last checking for robustness by implementing false rule changes in 2017 and 2019. The equations used to estimate these coefficients can be found in the Appendix.

Table 4: Regression output for Cap Number.

	Dependent variable:				
	Cap Number				
	(1)	(2)	(3)	(4)	(5)
QB	2,473,682***	10,638,205***	22,077,647***	2,302,324***	2,704,307***
	(140,746)	(243,089)	(893,608)	(157,684)	(129,098)
RI	59,342	433,328***	433,328***		
	(86,550)	(41,388)	(39,663)		
QB*RI	886,166***	2,318,999***	2,357,645**		
	(210,012)	(351,326)	(1,007,477)		
RI17				404,408***	
				(44,697)	
QB*RI17				1,014,540***	
				(210,620)	
RI19					451,911***
					(46,499)
QB*RI19					463,537**
					(220,162)
Constant	-171,205,740***	1,634,213***	1,634,213***	1,600,312***	1,677,836***
	(34,368,327)	(28,065)	(26,894)	(33,938)	(27,286)
Observations	21,372	20,704	20,465	21,372	21,372
Note:					*p**p***p<0.01

Note on the regressions: Standard errors are in parentheses. (1) Control for time trend, (2) Top 30 percent, (3) Top five percent, (4) Robust check for 2017 and (5) Robust check for 2019. In addition to this, we chose to leave out the variable “year” from the output in Table 4. Table 4 gives us an answer to central difficulties that arises when estimating DID. The planned salary for estimates (2) and (3) is more or less the same. This indicates that whether or not a QB is among the top 30 or top five percent highest paid players in the league, the rule

change only increased their salary with a little over 2.3 million dollars. However, as presumed when visualizing our time trend, we arrive at significant results from our robustness checks. If the rule change was implemented in 2017 (4), the increase in planned salaries for QBs would be around 950 thousand dollars, and a false implementation of the rule change in 2019 (5) would increase QBs planned salaries with around 500 thousand dollars. Though this is unfortunate, the speedy acceleration of planned QB salaries might explain this, and we will further discuss this in the discussion part of the paper.
We now do the same for Cash Spent:

Table 5: Regression output for Cash Spent.

	Dependent variable:				
	Cash Spent				
	(1)	(2)	(3)	(4)	(5)
QB	2,677,035***	11,890,572***	32,151,247***	2,687,692***	2,994,650***
	(168,167)	(293,740)	(791,998)	(188,428)	(154,230)
RI	177,682*	478,992***	478,992***		
	(103,413)	(49,846)	(47,145)		
QB*RI	1,199,318***	3,022,457***	1,330,012		
	(250,928)	(423,008)	(1,008,544)		
RI17				446,975***	
				(53,412)	
QB*RI17				941,754***	
				(251,685)	
RI19					488,437***
					(55,551)
QB*RI19					619,408**
					(263,022)
Constant	-137,479,468***	1,773,181***	1,773,181***	1,735,736***	1,825,222***
	(41,064,245)	(33,800)	(31,968)	(40,555)	(32,598)
Observations	21,372	20,704	20,465	21,372	21,372
Note:					*p**p***p<0.01

Note on the regressions: Standard errors are in parentheses. The notation (1:5) is the same for Cash Spent as the ones for Cap Number.
Table 5 show that for actual spending for the top 30 percent of players (2), we see that the increase now is around 3 million dollars, while the increase stemming from the rule change is

around 1.3 million dollars for the top five percent (3). Again, we end up with significant results from our robustness checks. This time it actually shows that the increase from a false rule change in 2017 (4) is around 1 million dollars, while implementing a false rule change in 2019 (5) gives an increase of 600 thousand dollars for QBs.

For time trends (1) in both occasions (Cap Number and Cash Spent), we see that the results for the δ-coefficient does not differ that much from the one we got without including the coefficient "year".

5. Discussion

What this paper sought out to do was ultimately to find out if there was any causality between the rule change and increasing salary for QBs. For our case, we found an increase in QB salaries when using 2018 as our intervention point.

The trouble with drawing inference from the results, are that our results do not account for an increasing Salary Cap. Our results only find an increase in actual salary, not relative salary. In addition to this, our results have a low adjusted R^2. This indicates that the scatter plot around our fitted regression line deviates a lot, but as we mentioned before, we did not set out to find variance in Cap Number or Cash Spent, and as we still find significant answers we are able to draw inference about our end results. Another unfortunate part of our research was found when doing our robustness checks and found a treatment effect, where there was no treatment (2017 and 2019). We find it reasonably to assume that this stems from the increasing acceleration of salaries for the Treatment group opposite to the slight increasing trend for the Control group.

A positive note on our results is that we see that the Pareto distribution still holds after the rule change. Both for top 30 and five percent we see a higher increase in salary after the rule change for all our results, but one. This furthermore proves that the top percent of QBs have an increasing salary after the rule change.

For researchers interested in this topic, a note to further research should be to control for an increasing Salary Cap. Controlling for Salary Cap would allow us to find answers about relative change in salaries stemming from the rule change. Additionally, the research would become more comprehensibly if there was implemented an econometric model for time series data with non-stationary data, as this would solve our problems with concluding that it is the rule change that increase the salaries for QBs.

6. Conclusion

We started this paper by noticing how the NFL has changed in terms of penalties called for defenders committing a foul leading to a RTP-penalty enforced. Furthermore, we introduced a little bit of previous literature and the existence of a Pareto distribution in regards to all players, QBs and the top 30- and five percent of QBs. From there we went on to explain how we collected the data used for estimation, before explaining the process of DID-estimation. Second to last we presented our positive results, which further proved that our hypotheses was true. We then chose to control for time trend, the top 30- and five percent of QBs, and lastly robustness checks for 2017 and 2019.

Going back to the research question:

- How did changing the "roughing the passer rule" in 2018 affect salaries for quarterbacks in the NFL?

Finding treatment effects where no intervention was implemented means we cannot decisively say that it was the rule change that drove QB salaries further up. However, what we can say is that the effect was largest in 2017, a little smaller in 2018, and even smaller in 2019.

7. References

Cahuc, P., Carcillo, S., & Zylberberg, A. (2014). *Labor economics* (Vol. 2end ed.). MIT Press.

Card, D., & Krueger, A. B. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply. *American Economic Review*, *90*(5), 1397-1420. https://doi.org/10.1257/aer.90.5.1397

Heubeck, T. S., J (2003). Incentive Clauses in Players' Contracts in Team Sports-Theory and Practice *German working papers in Law and Economics 2*.

Hill, R. C., Griffiths, W. E., & Lim, G. C. (2018). *Principles of econometrics* (5th ed. ed.). Wiley.

Mulholland, J., & Jensen, S. T. (2019). Optimizing the allocation of funds of an NFL team under the salary cap [Article]. *International Journal of Forecasting*, *35*(2), 767-775. https://doi.org/10.1016/j.ijforecast.2018.09.004

NFL Penalties. (2020). Nflpenalties.com. Retrieved 10.05.2021 from https://www.nflpenalties.com/penalty/roughing-the-passer?year=2020

NFLPA. (2020). CBA https://nflpaweb.blob.core.windows.net/website/PDFs/CBA/March-15-2020-NFL-NFLPA-Collective-Bargaining-Agreement-Final-Executed-Copy.pdf

Operations, F. (2018). *NFL Video Rulebook ROUGHING THE PASSER*. Nfl.com. Retrieved 14.04.2021 from https://operations.nfl.com/the-rules/nfl-video-rulebook/roughing-the-passer/

Pareto, V. (1896). LA CURVA DELLE ENTRATE E LE OSSERVAZIONI DEL PROF. EDGEWORTH. *Giornale degli Economisti*, *13 (Anno 7)*, 439-448. http://www.jstor.org/stable/23219510

PS. (2021). Overthecap.com. Retrieved 13.04.2021 from https://overthecap.com/position/

Roach, M. A. (2017). Testing Labor Market Efficiency Across Position Groups in the NFL. *Journal of Sports Economics*, *19*(8), 1093-1121. https://doi.org/10.1177/1527002517704021

Schalter, T. (2013). *Power Ranking the Imortance of Each Position in Today's NFL*. Bleacherrepport.com. Retrieved 15.04.2021 from https://bleacherreport.com/articles/1659834-power-ranking-the-importance-of-each-position-in-todays-nfl

8. Appendix

8.1 Main regression

Explanation of the variables:

$\mathbf{Y_{CapNumber}}$: Indicates how much money a team intends to spend on a player for a given year.

$\boldsymbol{Y_{CashSpent}}$: Indicates how much money a team actually spent on a player for a given year.

$\boldsymbol{QB}$: Dummy variable. Indicates whether we are looking at the Treatment- or Control group. (QB = 1 for the Treatment group, QB=0, for the Control group)

$\boldsymbol{RI}$: Dummy variable. Indicates whether we are before or after intervention. (RI = 1 after the rule change, RI = 0 before the rule change)

$\boldsymbol{\delta}$: The Difference-in-Differences estimator, explained by the Difference-in-Differences matrix below.

The two regressions:

$$Y_{CapNumber} = \beta_0 + \beta_1 QB + \beta_2 RI + \delta(QB * RI) + \epsilon \quad (1)$$

$$Y_{CashSpent} = \beta_0 + \beta_1 QB + \beta_2 RI + \delta(QB * RI) + \epsilon \quad (2)$$

Difference-in-Differences matrix:

	QB = 1	QB = 0	Change
R-I = 1	$\beta_0 + \beta_1 + \beta_2 + \delta$	$\beta_0 + \beta_2$	$\beta_1 + \delta$
R-I = 0	$\beta_0 + \beta_1$	β_0	β_1
Change	$\beta_2 + \delta$	β_2	δ

How we end up with $\boldsymbol{\delta}$ *being the coefficient that determines Change:*

First proof:	**Second proof:**
$[(\beta_0+\beta_1+\beta_2+\delta)-(\beta_0+\beta_1)] - [(\beta_0+\beta_2)-(\beta_0)]$	$[(\beta_0+\beta_1+\beta_2+\delta)-(\beta_0+\beta_2)] - [(\beta_0+\beta_1)-(\beta_0)]$
$= [\beta_0-\beta_0+\beta_1-\beta_1+\beta_2+\delta] - [\beta_0-\beta_0+\beta_2]$	$= [\beta_0-\beta_0+\beta_1+\beta_2-\beta_2+\delta] - [\beta_0-\beta_0+\beta_1]$
$= [\beta_2+\delta-\beta_2] = [\beta_2-\beta_2+\delta]$	$= [\beta_1+\delta-\beta_1] = [\beta_1-\beta_1+\delta]$
$= \boldsymbol{\delta}$	$= \boldsymbol{\delta}$

8.2 Control regressions

Explanation of the variables:

$\boldsymbol{Year}$ = A numerical variable, indicating what year the observation represents. Stretching from 2013-2021. Since we use δ as our β_3 value, $Year$ will be noted with β_4.

$\boldsymbol{TOP30QB}$ = Dummy variable. Created new sample only including the top 30 percent of QBs. ($TOP30QB$ = 1 for the top 30 percent of the Treatment group, $TOP30QB$ = 0 for the Control group)

$\boldsymbol{TOP5QB}$ = Dummy variable. Created new sample only including the top five percent of QBs. ($TOP5QB$ = 1 for the top five percent of the Treatment group, $TOP5QB$ = 0 for the Control group)

$\boldsymbol{RI17}$ = Dummy variable. Indicating whether we are before or after the false intervention in 2017. ($RI17$ = 1 after, $RI17$ = 0 before).

$\boldsymbol{RI19}$ = Dummy variable. Indicating whether we are before or after false intervention in 2019. ($RI19$ = 1 after, $RI19$ = 0 before).

The five regressions:

Cap Number:

$$\boldsymbol{Y_{CapNumber}} = \beta_0 + \beta_1 QB + \beta_2 RI + \delta(QB * RI) + \beta_4 Year + \epsilon \tag{5}$$

$$\boldsymbol{Y_{CapNumber}} = \beta_0 + \beta_1 TOP30QB + \beta_2 RI + \delta(TOP30QB * RI) + \epsilon \tag{6}$$

$$\boldsymbol{Y_{CapNumber}} = \beta_0 + \beta_1 TOP5QB + \beta_2 RI + \delta(TOP5QB * RI) + \epsilon \tag{7}$$

$$\boldsymbol{Y_{CapNumber}} = \beta_0 + \beta_1 QB + \beta_2 RI17 + \delta(QB * RI17) + \epsilon \tag{8}$$

$$\boldsymbol{Y_{CapNumber}} = \beta_0 + \beta_1 QB + \beta_2 RI19 + \delta(QB * RI19) + \epsilon \tag{9}$$

Cash Spent:

$$\boldsymbol{Y_{CashSpent}} = \beta_0 + \beta_1 QB + \beta_2 RI + \delta(QB * RI) + \beta_4 Year + \epsilon \tag{10}$$

$$\boldsymbol{Y_{CashSpent}} = \beta_0 + \beta_1 TOP30QB + \beta_2 RI + \delta(TOP30QB * RI) + \epsilon \tag{11}$$

$$\boldsymbol{Y_{CashSpent}} = \beta_0 + \beta_1 TOP5QB + \beta_2 RI + \delta(TOP5QB * RI) + \epsilon \tag{12}$$

$$\boldsymbol{Y_{CashSpent}} = \beta_0 + \beta_1 QB + \beta_2 RI17 + \delta(QB * RI17) + \epsilon \tag{13}$$

$$\boldsymbol{Y_{CashSpent}} = \beta_0 + \beta_1 QB + \beta_2 RI19 + \delta(QB * RI19) + \epsilon \tag{14}$$

8.3 R-codes

```
# Loading the packages needed
library(readr)
library(tidyverse)
library(stats)
library(stargazer)
library(dplyr)
library(mosaic)
library(graphics)

# DID-estimation, using player specific data
# Cleaning the environment
rm(list=ls())

# In the scrape section later on,
# you will see that we saved the finished
# data set to our system.
# However, we wanted it to be easier for you
# to extract the data,
# so here is an exact replica of our data-set:
nfltotal<-read_csv("https://raw.githubusercontent.com/thojak14/MasterThesis-/main/NFLPS.csv")
nfltotal$X1 <- NULL

# Creating the dummy called "rule_implemented"
nfltotal$rule_implemented=ifelse(nfltotal$year>=2018,1,0)

# Correlation
cncorr<-cor.test(nfltotal$qb,nfltotal$cap_number,
                    method = "pearson")
cncorr # 0.18<0.29=low degree of correlation

cscorr<-cor.test(nfltotal$qb,nfltotal$cash_spent,
                   method = "pearson")
cscorr #0.17<0.29=low degree of correlation

# Descriptive statistics
nfltotaldataframe <- data.frame(nfltotal)
stargazer(nfltotaldataframe, digits=0,
          summary.stat=c("n","mean","sd","max"),
          keep=c("cap_number","cash_spent",
                 "salary_cap"),
          covariate.labels=c("Cap Number","Cash Spent",
                             "Salary Cap"),
          title="Descriptive Statistics",type="html",out="DS.doc")

# Creating a time trend of the raw data.
# We do this by creating mean values for each year,
```

```
# to both Cap Number and Cash Spent. This is done
# for both the Treatment group and the Control group.

# Treatment group
qb2013 <- nfltotal %>% filter(qb==1) %>% filter(year==2013)
qb2014 <- nfltotal %>% filter(qb==1) %>% filter(year==2014)
qb2015 <- nfltotal %>% filter(qb==1) %>% filter(year==2015)
qb2016 <- nfltotal %>% filter(qb==1) %>% filter(year==2016)
qb2017 <- nfltotal %>% filter(qb==1) %>% filter(year==2017)
qb2018 <- nfltotal %>% filter(qb==1) %>% filter(year==2018)
qb2019 <- nfltotal %>% filter(qb==1) %>% filter(year==2019)
qb2020 <- nfltotal %>% filter(qb==1) %>% filter(year==2020)
qb2021 <- nfltotal %>% filter(qb==1) %>% filter(year==2021)

# Control group
nonqb2013 <- nfltotal %>% filter(qb==0) %>% filter(year==2013)
nonqb2014 <- nfltotal %>% filter(qb==0) %>% filter(year==2014)
nonqb2015 <- nfltotal %>% filter(qb==0) %>% filter(year==2015)
nonqb2016 <- nfltotal %>% filter(qb==0) %>% filter(year==2016)
nonqb2017 <- nfltotal %>% filter(qb==0) %>% filter(year==2017)
nonqb2018 <- nfltotal %>% filter(qb==0) %>% filter(year==2018)
nonqb2019 <- nfltotal %>% filter(qb==0) %>% filter(year==2019)
nonqb2020 <- nfltotal %>% filter(qb==0) %>% filter(year==2020)
nonqb2021 <- nfltotal %>% filter(qb==0) %>% filter(year==2021)

# Creating data frame of mean values for Cap Number
# Treatment group
cnmeanqb <- data.frame(mean(qb2013$cap_number),
                       mean(qb2014$cap_number),
                       mean(qb2015$cap_number),
                       mean(qb2016$cap_number),
                       mean(qb2017$cap_number),
                       mean(qb2018$cap_number),
                       mean(qb2019$cap_number),
                       mean(qb2020$cap_number),
                       mean(qb2021$cap_number))

# Control group
cnmeannonqb <- data.frame(mean(nonqb2013$cap_number),
                          mean(nonqb2014$cap_number),
                          mean(nonqb2015$cap_number),
                          mean(nonqb2016$cap_number),
                          mean(nonqb2017$cap_number),
                          mean(nonqb2018$cap_number),
                          mean(nonqb2019$cap_number),
                          mean(nonqb2020$cap_number),
                          mean(nonqb2021$cap_number))

# Creating data frame of mean values for Cash Spent
```

```
csmeanqb <- data.frame(mean(qb2013$cash_spent),
                       mean(qb2014$cash_spent),
                       mean(qb2015$cash_spent),
                       mean(qb2016$cash_spent),
                       mean(qb2017$cash_spent),
                       mean(qb2018$cash_spent),
                       mean(qb2019$cash_spent),
                       mean(qb2020$cash_spent),
                       mean(qb2021$cash_spent))

# Control group
csmeannonqb <- data.frame(mean(nonqb2013$cash_spent),
                          mean(nonqb2014$cash_spent),
                          mean(nonqb2015$cash_spent),
                          mean(nonqb2016$cash_spent),
                          mean(nonqb2017$cash_spent),
                          mean(nonqb2018$cash_spent),
                          mean(nonqb2019$cash_spent),
                          mean(nonqb2020$cash_spent),
                          mean(nonqb2021$cash_spent))

# Transpose to easier plot the results
cnmeanqb <- t(cnmeanqb)
cnmeannonqb <- t(cnmeannonqb)
csmeanqb <- t(csmeanqb)
csmeannonqb <- t(csmeannonqb)

# Plot
plot(cnmeanqb/1000000,xaxt="n",ylim=c(0,7),
     main="Time trend Treatment group",
     xlab="Year",ylab="Cap Number",type="b")
axis(1,at=1:9,labels=c(2013:2021))

plot(cnmeannonqb/1000000, xaxt="n",ylim=c(0,7),
     main="Time trend Control group",
     xlab="Year",ylab="Cap Number",type="b")
axis(1,at=1:9,labels=c(2013:2021))

plot(csmeanqb/1000000,xaxt="n",ylim=c(0,7),
     main="Time trend Treatment group",
     xlab="Year",ylab="Cash Spent",type="b")
axis(1,at=1:9,labels=c(2013:2021))

plot(csmeannonqb/1000000, xaxt="n",ylim=c(0,7),
     main="Time trend Control group",
     xlab="Year",ylab="Cash Spent",type="b")
axis(1,at=1:9,labels=c(2013:2021))
```

```
# Estimating DID using our regressions
didqbcn <- lm(cap_number~qb+rule_implemented+
                qb*rule_implemented, data=nfltotal)
didcsqb <- lm(cash_spent~qb+rule_implemented+
                qb*rule_implemented, data=nfltotal)

# Results
summary(didqbcn) # delta-coefficient=897,399
summary(didcsqb) # delta-coefficient=1,208,367

# Creating table for the article
stargazer(didqbcn,didcsqb, digits=0,
          dep.var.labels=c("Cap Number","Cash Spent"),
          covariate.labels=c("QB","RI","QB*RI"),
          keep.stat=c("n","adj.rsq","f"),
          title="DID",type="html",out="DID.doc")

# Creating before treatment, after treatment,
# before control and after control data sets.
treatafter <- nfltotal %>% filter(qb==1) %>%
  filter(rule_implemented==1)
treatbefore <- nfltotal %>% filter(qb==1) %>%
  filter(rule_implemented==0)
controlafter <- nfltotal %>% filter(qb==0) %>%
  filter(rule_implemented==1)
controlbefore <- nfltotal %>% filter(qb==0) %>%
  filter(rule_implemented==0)

# Creating data frames consisting
# of mean values.
didcn <- data.frame(mean(treatafter$cap_number),
                    mean(treatbefore$cap_number),
                    mean(controlafter$cap_number),
                    mean(controlbefore$cap_number))
didcs <- data.frame(mean(treatafter$cash_spent),
                    mean(treatbefore$cash_spent),
                    mean(controlafter$cash_spent),
                    mean(controlbefore$cash_spent))

# Finding the delta using mean values
round((didcn[1]-didcn[3])-(didcn[2]-didcn[4]),0) ->
  DeltaCN # 897,399
round((didcs[1]-didcs[3])-(didcs[2]-didcs[4]),0) ->
  DeltaCS  # 1,208,367

# Adding the delta to mean values
meanvaluescn <- cbind(didcn,DeltaCN)
meanvaluescs <- cbind(didcs,DeltaCS)
```

```
# Giving the values new names
names(meanvaluescn)[1:5] <- c("Treatment After","Treatment Before",
                              "Control After","Control After","Delta
")
names(meanvaluescs)[1:5] <- c("Treatment After","Treatment Before",
                              "Control After","Control After","Delta
")

# Creating tables for article
stargazer(meanvaluescn,digits=0,
          summary.stat="mean",out="MeanCN.doc",
          title="Mean values Cap Number",type="html")
stargazer(meanvaluescs,digits=0,
          summary.stat="mean",out="MeanCS.doc",
          title="Mean values Cash Spent",type="html")

# Creating plot for DID.
# Note: The following code was more or less copied from "Chapter 7"
# in Principles of Econometrics (R.Carter Hill * William E. Griffith
s * Guay C.Lim).
# Only numbers concerning our research was put in place of the code
copied from Chapter 7.
# We also chose to devide the numbers with 1 million.
# This because we wanted to show DID in millions of dollars.

# Cap number.
b0 <- coef(didqbcn)[1]/1000000
b1 <- coef(didqbcn)[2]/1000000
b2 <- coef(didqbcn)[3]/1000000
delta <- coef(didqbcn)[4]/1000000

# Creating TA,CA,TB,CB,CTA
TA <- b0+b1+b2+delta
CA <- b0+b2
TB <- b0+b1
CB <- b0
CTA <- CA+(TB-CB)

# Plot.
# Note: The whole section needs to be run at the same time.
plot(1,type="n",xlab="Intervention",ylab="Cap Number",
     xaxt="n",xlim=c(-0.01,1.01),ylim=c(1.5, 6),
     main="Difference-in-Differences")
segments(x0=0,y0=CB,x1=1,y1=CA,lty=1,col=2,lwd=2)
segments(x0=0,y0=TB,x1=1,y1=TA,lty=1,col=4,lwd=2)
segments(x0=0,y0=TB,x1=1,y1=CTA,lty=2,col=3,lwd=2)
legend("topleft",
       legend=c("Control","Treatment",
                "Counterfactual Treatment"),
```

```r
       lty=c(1,1,2),col=c(2,4,3))
axis(side=1,at=c(0,1),labels=NULL)
axis(side=2,seq(1,6,by=0.5))
text(0,1.73,"CB");text(0,4.20,"TB");text(1,5.53,"TA");
text(1,4.63,"CTA");text(1,2.17,"CA")

# Hence we have proved that:
round(((TA-CTA)*1000000),0) # delta-coefficient=897,399

# Cash spent.
b0 <- coef(didcsqb)[1]/1000000
b1 <- coef(didcsqb)[2]/1000000
b2 <- coef(didcsqb)[3]/1000000
delta <- coef(didcsqb)[4]/1000000

# Creating TA,CA,TB,CB,CTA
TA <- b0+b1+b2+delta
CA <- b0+b2
TB <- b0+b1
CB <- b0
CTA <- CA+(TB-CB)

# Plot
plot(1,type="n",xlab="Intervention",ylab="Cash Spent",
     xaxt="n",xlim=c(-0.01,1.01),ylim=c(1.5,6.5),
     main="Difference-in-Differences")
segments(x0=0,y0=CB,x1=1,y1=CA,lty=1,col=2,lwd=2)
segments(x0=0,y0=TB,x1=1,y1=TA,lty=1,col=4,lwd=2)
segments(x0=0,y0=TB,x1=1,y1=CTA,lty=2,col=3,lwd=2)
legend("topleft",legend=c("Control","Treatment",
                          "Counterfactual Treatment"),
       lty=c(1,1,2),col=c(2,4,3))
axis(side=1,at=c(0,1),labels=NULL)
axis(side=2,seq(1,6.5,by=0.5))
text(0,1.87,"CB");text(0,4.54,"TB");text(1,6.23,"TA");
text(1,5.02,"CTA");text(1,2.35,"CA")

# Hence we have proved that:
round(((TA-CTA)*1000000),0)# delta-coefficient=1,208,367

# DID with time trend included
didtimecn <- lm(cap_number~qb+rule_implemented+
                  qb*rule_implemented+year, data=nfltotal)
didtimecs <- lm(cash_spent~qb+rule_implemented+
                  qb*rule_implemented+year, data=nfltotal)

# Results
summary(didtimecn)
summary(didtimecs)
```

```
# DID with Robustness checks

# Assuming that the rule is changed in 2017 or 2019
nfltotal$rule_implemented17=ifelse(nfltotal$year>=2017,1,0)
nfltotal$rule_implemented19=ifelse(nfltotal$year>=2019,1,0)

# Estimating DID
didcnqb17 <- lm(cap_number~qb+rule_implemented17+
                  qb*rule_implemented17,data=nfltotal)
didcsqb17 <- lm(cash_spent~qb+rule_implemented17+
                  qb*rule_implemented17,data=nfltotal)
didcnqb19 <- lm(cap_number~qb+rule_implemented19+
                  qb*rule_implemented19, data=nfltotal)
didcsqb19 <- lm(cash_spent~qb+rule_implemented+
                  qb*rule_implemented19, data=nfltotal)

# Results
summary(didcnqb17)
summary(didcsqb17)
summary(didcnqb19)
summary(didcsqb19)

# Estimating DID for the top percentage
# of QBs.

# Creating subsets for QBs and non QBs
nfltotalqb <- subset(nfltotal, qb==1)
nfltotalnonqb <- subset(nfltotal, qb==0)

# Creating a decending order for all players
nfltotal[rev(order(nfltotal$cap_number)),] %>%
  head(21372) -> nfltotaltopcn
nfltotal[rev(order(nfltotal$cash_spent)),] %>%
  head(21372) -> nfltotaltopcs

# Then, creating a descending order for QBs
nfltotalqb[rev(order(nfltotalqb$cap_number)),] %>%
  head(954) -> top_cap_number_qb
nfltotalqb[rev(order(nfltotalqb$cash_spent)),] %>%
  head(954) -> top_cash_spent_qb

# 30 % = approximately the top 286 observations
nfltotalqb[rev(order(nfltotalqb$cap_number)),] %>%
  head(286) -> top_cap_number_qb_30
nfltotalqb[rev(order(nfltotalqb$cash_spent)),] %>%
  head(286) -> top_cash_spent_qb_30

# 5 % = approximately the top 47 observations
nfltotalqb[rev(order(nfltotalqb$cap_number)),] %>%
```

```
  head(47) -> top_cap_number_qb_5
nfltotalqb[rev(order(nfltotalqb$cash_spent)),] %>%
  head(47) -> top_cash_spent_qb_5

# Creating different datasets
# One for top 30 % of QBs and all the others,
# And one for top 5 % of QBs and all the others.
nfltotalcn30 <- rbind(nfltotalnonqb,top_cap_number_qb_30)
nfltotalcs30 <- rbind(nfltotalnonqb,top_cash_spent_qb_30)
nfltotalcn5 <- rbind(nfltotalnonqb,top_cap_number_qb_5)
nfltotalcs5 <- rbind(nfltotalnonqb,top_cash_spent_qb_5)

# DID top 30 percent of QBs
didtop30cn <- lm(cap_number~qb+rule_implemented+
                   qb*rule_implemented, data=nfltotalcn30)
didtop30cs <- lm(cash_spent~qb+rule_implemented+
                   qb*rule_implemented, data=nfltotalcs30)

# DID top five percent of QBs
didtop5cn <- lm(cap_number~qb+rule_implemented+
                  qb*rule_implemented, data=nfltotalcn5)
didtop5cs <- lm(cash_spent~qb+rule_implemented+
                  qb*rule_implemented, data=nfltotalcs5)

# Results
summary(didtop30cn)
summary(didtop30cs)
summary(didtop5cn)
summary(didtop5cs)

# Now that we have created subsets consisting of
# the top 30 and five percent of QBs,
# we are able to plot graphs for the article.

# Plot all players
plot(nfltotaltopcn$cap_number/1000000,
     main="Projected distribution of money for all players",
     xlab="Players",
     ylab="Cap Number")
plot(nfltotaltopcs$cash_spent/1000000,
     main="Actual distribution of money for all players",
     xlab="Players",
     ylab="Cash Spent")

# Plot all QBs
plot(top_cap_number_qb$cap_number/1000000,
     main="Projected distribution of money for QBs",
     xlab="QBs",
     ylab="Cap Number")
```

```r
plot(top_cash_spent_qb$cash_spent/1000000,
     main="Actual distribution of money for QBs",
     xlab="QBs",
     ylab="Cash Spent")

# Plot top 30 percent of QBs
plot(top_cap_number_qb_30$cap_number/1000000,
     main="Projected distribution of money for top 30 percent of QBs
",
     xlab="QBs",
     ylab="Cap Number")
plot(top_cash_spent_qb_30$cash_spent/1000000,
     main="Actual distribution of money for top 30 percent of QBs",
     xlab="QBs",
     ylab="Cash Spent")

# Plot top 5 percent of QBs
plot(top_cap_number_qb_5$cap_number/1000000,
     main="Projected distribution of money for top five percent of Q
Bs",
     xlab="QBs",
     ylab="Cap Number")
plot(top_cash_spent_qb_5$cash_spent/1000000,
     main="Actual distribution of money for top five percent of QBs"
,
     xlab="QBs",
     ylab="Cash Spent")

# Creating two tables that consists of
# Control for time trends,
# Robustness checks and
# Top 30 and five percent of QBs.
# One for Cap Number and one for
# Cash Spent.
# We omit "year" as we are only interested in the
# interaction variable between "qb" and "rule implemented",
# the interaction between "qb30" & "qb5" and "rule implemented
# and the interaction between "qb" and "rule implemented 17" &
# "rule implemented 19".


# Cap Number
stargazer(didtimecn,didtop30cn,didtop5cn,
          didcnqb17,didcnqb19, digits=0,
          dep.var.labels= "Cap Number",
          covariate.labels = c("QB","RI","QB*RI","RI17",
                               "QB*RI17","R-I19","QB*RI19"),
          keep.stat="n",
          type="html",out="capnumber.doc",omit="year")
```

```r
# Cash Spent
stargazer(didtimecs,didtop30cs,didtop5cs,
          didcsqb17,didcsqb19, digits=0,
          dep.var.labels = "Cash Spent",
          covariate.labels = c("QB","RI","QB*RI","RI17",
                               "QB*RI17","RI19","QB*RI19"),
          keep.stat="n",
          type="html",out="cashspent.doc",omit="year")

# Scrape for player specific data
# Note: Scraping player specific data consisted of scraping
# each position for each year. However, in this Appendix
# we will only include the scrape for the data
# concerning 2021.

# Loading packages
library(tidyverse)
library(xml2)
library(rvest)
library(janitor)
library(dplyr)

# Cleaning the environment
rm(list=ls())

# First, scraping information about each position.
# Then, cleaning the column names, so we do not use capital letters.
# Last, we create a variable for "year" and a dummy for QB.

# 2021

# Offense
qb2021 <- read_html("https://overthecap.com/position/quarterback/202
1/")
%>% html_table(fill=TRUE)
qb2021 <- qb2021[[1]]
qb2021 <- qb2021 %>% clean_names()
qb2021$year <-(2021)
qb2021$qb <- (1)
rb2021 <- read_html("https://overthecap.com/position/running-back/20
21/")
%>% html_table(fill=TRUE)
rb2021 <- rb2021[[1]]
rb2021$year <-(2021)
rb2021 <- rb2021 %>% clean_names()
rb2021$qb <-(0)
fb2021 <- read_html("https://overthecap.com/position/fullback/2021/"
)
```

```
%>% html_table(fill=TRUE)
fb2021 <- fb2021[[1]]
fb2021$year <-(2021)
fb2021 <-fb2021 %>% clean_names()
fb2021$qb <-(0)
wr2021 <- read_html("https://overthecap.com/position/wide-receiver/2
021/")
%>% html_table(fill=TRUE)
wr2021 <- wr2021[[1]]
wr2021$year <-(2021)
wr2021 <- wr2021 %>% clean_names()
wr2021$qb <-(0)
te2021 <- read_html("https://overthecap.com/position/tight-end/2021/
")
%>% html_table(fill=TRUE)
te2021 <- te2021[[1]]
te2021$year <-(2021)
te2021 <- te2021 %>% clean_names()
te2021$qb <-(0)
lt2021 <- read_html("https://overthecap.com/position/left-tackle/202
1/")
%>% html_table(fill=TRUE)
lt2021 <- lt2021[[1]]
lt2021$year <-(2021)
lt2021 <- lt2021 %>% clean_names()
lt2021$qb <-(0)
lg2021 <- read_html("https://overthecap.com/position/left-guard/2021
/")
%>% html_table(fill=TRUE)
lg2021 <- lg2021[[1]]
lg2021$year <-(2021)
lg2021 <- lg2021 %>% clean_names()
lg2021$qb <-(0)
ce2021 <- read_html("https://overthecap.com/position/center/2021/")
%>% html_table(fill=TRUE)
ce2021 <- ce2021[[1]]
ce2021$year <-(2021)
ce2021 <- ce2021 %>% clean_names()
ce2021$qb <-(0)
rg2021 <- read_html("https://overthecap.com/position/right-guard/202
1/")
%>% html_table(fill=TRUE)
rg2021 <- rg2021[[1]]
rg2021$year <-(2021)
rg2021 <- rg2021 %>% clean_names()
rg2021$qb <-(0)
rt2021 <- read_html("https://overthecap.com/position/right-tackle/20
21/")
%>% html_table(fill=TRUE)
```

```r
rt2021 <- rt2021[[1]]
rt2021$year <-(2021)
rt2021 <-rt2021 %>% clean_names()
rt2021$qb <-(0)

# Defense
idl2021 <- read_html("https://overthecap.com/position/interior-defen
sive-line/2021/")
%>% html_table(fill=TRUE)
idl2021 <- idl2021[[1]]
idl2021$year <-(2021)
idl2021 <- idl2021 %>% clean_names()
idl2021$qb <-(0)
ed2021 <- read_html("https://overthecap.com/position/edge-rusher/202
1/")
%>% html_table(fill=TRUE)
ed2021 <- ed2021[[1]]
ed2021$year <-(2021)
ed2021 <- ed2021 %>% clean_names()
ed2021$qb <-(0)
lb2021 <- read_html("https://overthecap.com/position/linebacker/2021
/")
%>% html_table(fill=TRUE)
lb2021 <- lb2021[[1]]
lb2021$year <-(2021)
lb2021 <- lb2021 %>% clean_names()
lb2021$qb <-(0)
sa2021 <- read_html("https://overthecap.com/position/safety/2021/")
%>% html_table(fill=TRUE)
sa2021 <- sa2021[[1]]
sa2021$year <-(2021)
sa2021 <- sa2021 %>% clean_names()
sa2021$qb <-(0)
cb2021 <- read_html("https://overthecap.com/position/cornerback/2021
/")
%>% html_table(fill=TRUE)
cb2021 <- cb2021[[1]]
cb2021$year <-(2021)
cb2021 <- cb2021 %>% clean_names()
cb2021$qb <-(0)

# Special teams
ki2021 <- read_html("https://overthecap.com/position/kicker/2021/")
%>% html_table(fill=TRUE)
ki2021 <- ki2021[[1]]
ki2021$year <- (2021)
ki2021 <- ki2021 %>% clean_names()
ki2021$qb <-(0)
pu2021 <- read_html("https://overthecap.com/position/punter/2021/")
```

```
%>% html_table(fill=TRUE)
pu2021 <- pu2021[[1]]
pu2021$year <-(2021)
pu2021 <- pu2021 %>% clean_names()
pu2021$qb <-(0)
ls2021 <- read_html("https://overthecap.com/position/long-snapper/20
21/")
%>% html_table(fill=TRUE)
ls2021 <- ls2021[[1]]
ls2021$year <-(2021)
ls2021 <- ls2021 %>% clean_names()
ls2021$qb <-(0)

# Creating a new data frame consisting of all the positions
players2021 <- rbind(qb2021,cb2021,ce2021,ed2021,fb2021,
                     idl2021,ki2021,lb2021,lg2021,ls2021,
                     lt2021,pu2021,rb2021,rg2021,rt2021,
                     sa2021,te2021,wr2021)

# Removing dollar signs and commas from cap_number and cash_spent
players2021[] <- lapply(players2021,gsub,pattern="$",
                        fixed=TRUE,replacement="")
players2021[] <- lapply(players2021,gsub,pattern=",",
                        fixed=TRUE,replacement="")

# Saving the data
save(players2021, file="players2021")

# Loading data for 2013-2021
load("players2013")
load("players2014")
load("players2015")
load("players2016")
load("players2017")
load("players2018")
load("players2019")
load("players2020")
load("players2021")

# Creating a column for salary-cap for each year, which will be our
dependent variable

# 2013,
# https://www.nfl.com/news/nfl-sets-2013-salary-cap-at-123m-up-from-
120-6m-0ap1000000146046
players2013$salary_cap <- (123000000)
players2013 <- players2013 %>%
  mutate_at(c(3:7), as.numeric)
str(players2013)
```

```r
# 2014,
# https://www.nfl.com/news/nfl-salary-cap-makes-nearly-10m-jump-to-1
33-million-0ap2000000329753
players2014$salary_cap <- (133000000)
players2014 <- players2014 %>%
  mutate_at(c(3:7), as.numeric)
str(players2014)

# 2015,
# https://www.nfl.com/news/nfl-salary-cap-will-be-143-28-million-in-
2015-0ap3000000475775
players2015$salary_cap <- (143280000)
players2015 <- players2015 %>%
  mutate_at(c(3:7), as.numeric)
str(players2015)

# 2016,
# https://nflpa.com/press/2016-adjusted-team-salary-caps
# Each team has individual, but using a league wide cap seems sufici
ent
players2016$salary_cap <- (155270000)
players2016 <- players2016 %>%
  mutate_at(c(3:7), as.numeric)
str(players2016)

# 2017,
# Reported to be 167.million
players2017$salary_cap <- (167000000)
players2017 <- players2017 %>%
  mutate_at(c(3:7), as.numeric)
str(players2017)

# 2018,
# Reported to be 177,2 million
players2018$salary_cap <- (177200000)
players2018 <- players2018 %>%
  mutate_at(c(3:7), as.numeric)
str(players2018)

# 2019
# Reported to be 188,2 million
players2019$salary_cap <- (188200000)
players2019 <- players2019 %>%
  mutate_at(c(3:7), as.numeric)
str(players2019)

# 2020,
# https://www.nfl.com/news/building-the-best-nfl-team-money-can-buy-
```

```r
under-the-2020-salary-cap
players2020$salary_cap <- (198200000)
players2020 <- players2020 %>%
  mutate_at(c(3:7), as.numeric)
str(players2020)


# 2021,
# https://www.nfl.com/news/nfl-2021-salary-cap-182-5-million
players2021$salary_cap <- (182500000)
players2021 <- players2021 %>%
  mutate_at(c(3:7), as.numeric)
str(players2021)

# Creating new data frame
nfltotal <- rbind(players2013,players2014,players2015,players2016,pl
ayers2017,
                  players2018,players2019,players2020,players2021)

# Saving the data frame, and then loading it again when estimating
save(nfltotal, file="NFLPS")

# Date for final scrape: 13th of April 2021, 13.00 PM.
```