EMORY UNIVERSITY
DEPARTMENT OF MATHEMATICS

Atlanta (GA) USA

Project 2

MATH345 MATH MODELING

Authors:

Amani Dabriwala

Carrie Gu

Hung Ngo

Tianyang Hu

David Zhang

Submitted By:
Group Uranus

Spring 2022

Group Report 2
today

EMORY
UNIVERSITY

Assignment 2: Formal Justice

The concept of formal justice or meritocracy relates to the idea that there should be a relationship between the qualifications of professionals and the compensation received for their work. These qualifications contain positive criteria (level of education, skill, and seniority), and negative criteria as well (the number and severity of mistakes committed on the job, etc.) An individual's compensation should bear a reasonable relationship to these qualifications in order to establish a fair wage scale. In this project you are requested to design a model relating compensation to the qualification(s). Assuming that one qualification could be ordered on a scale, the compensation (e.g., in $) should be "proportional" to it. This is well known since Aristotle times. Design a model to include this concept of meritocracy or formal justice in your compensation system. Start with only one qualification. Test the consistency of your model against the following data.

Company 1

	Qualif.: Seniority	Compensation (Yearly)
Empl. A	25 yrs	500,000 $
Empl. B	16 yrs	60,000 $
Empl. C	3 yrs	40,000 $

Company 2: NFL in 2003

Qualif.: Seniority	Compensation (Yearly)
Rookie	250,000 $
2nd y	300,000 $
3rd y	375,000 $
4th-6th y	450,000 $
7th-9th y	655,000 $
10th+ y	755,000 $

Is you model consistent with these data? How would you adapt your model to make it consistent? Are all the employers/players uniformly considered?

In a second stage of your project, consider a multiple qualification model. Extend your previous model to two (or more qualifications) and discuss how it works on two benchmarks.

Academic salaries The additional qualification beyond seniority is the education. Rank the education based on the Ph.D. institution of the professor and propose a model for the salary based on the two qualifications.

NFL Bonus Add to the previous model the qualification due to the results of the previous season. Change the salary model based on the achievement of some results (e.g. ranking in the conference,... to Superbowl win).

Discuss the quality of your model and how it can be adapted to negative qualifications (e.g. number of intercepts for a Quarterback over the previous season, missed field goals for a kicker, etc.)

Contribution

Amani Dabriwala
-presentation

Carrie Gu
-Idea Generation Process
-Introduction
-Assumption
-Data Collection

Tianyang Hu
-Idea Generation Process
- Assumption
-3.3.2 Models for NFL Players
-4.3 Future work

Hung Ngo
-Idea Generation Process
-3.1 Variables
-3.3.1 Models for Professors
-4.1 Strength

David Zhang
-Idea Generation Process
-3.2 One Qualification Model
-4.2 Weakness
-Abstract

Abstract

In stage 1, two models regarding the relationship between seniority and compensation are created. For company 1, the model uses the approach of polynomial interpolation, while for company 2, the model applies the least square fitting.

In stage 2, the first model is created for academic workers. Based on the result in stage 1, a linear relationship between seniority and compensation is assumed, and gender and education are included in the expanded model.

The second model is designed to understand the compensation allocation for NFL players. It understands the compensation for a specific player as the multiple of his share in the club and the club's total revenue. By further expanding the share and total revenue, the model considers factors like prize value for the team, probability of winning, efforts of each player, and seniority.

1 Introduction

Earnings inequality has played a major role in scholarly investigation since monetary resources is the foundation for other resources, such as power, authority, and autonomy. [11] It's important to accurately compensating individuals for their professional qualifications, which is known as formal justice. Sociologists have examined various factors that contributes to earning inequality, such as time [8], place [5], industry [6], and occupation [13]. Studies have also investigated the underlying effect of contextual factors on salary, such as race and gender. As the proportion of women and minorities in the labor market increased in the past decades, researchers began to consider the relationship between contextual factors and salary and found that female in academia generally received 16% less than male workers [2] [7].

Researchers have applied machine learning techniques and regression models to predict salary in the past. Navyashree et al. [9] conducted a salary prediction study in IT with several data mining techniques such as random forest, decision tree and support vector machine. Another study used linear regression and polynomial rergession to predict salary price given their position title [4]. However, those regression techniques require a large amount of data and may not be applicable when the data is limited. Therefore, in this project, we would like to explore the factors that determines an individual's salary by building our own model with a series of either positive or negative variables. The first stage of the model will only consider seniority as a factor of salary. The model will be further elaborated for two population group: NFL athletes and academic workers, with more variety of factors.

2 Assumption

For the model to be accurate, some assumptions are made to hold its validity:

- The individuals are living and working in the U.S. and follow the salary pattern in the U.S.
- The professor's salary is assumed to be affected by only three qualifications: seniority, gender, and education.
- The institution ranking on U.S. News is assumed to be the standard, and every institution consults it.

- The difference in average male and female professors' salaries reported in 2019 is assumed to stay the same.
- The players in NFL are both rational and risk neutral. That is they are indifferent to take risks in training as long as there are potential paybacks from their efforts.
- The club is assumed to publicize its value to the public to receive bonus from sponsors purely via advertisements on their player's training efforts and past prizes.
- The share of member on the club's revenue is linearly correlated with the member's seniority and training performance.

3 Formula

3.1 Variables

Variable	Definition	Unit
Sen	Seniority: Number of years of experience.	years
Edu	Education: The rank of the Ph.D institution	
Com	Compensation: Amount of paid salary each year	\$
C	Base Salary: The minimum salary	\$
S_i	The i-th player's salary	\$
r_i	The share of player i on the club revenues	
s	The club's revenue	\$
e_i	The training efforts for the i-th player	\$
V_{past}	The past prize values	\$
V	The potential prize values	\$
c_i	The costs of efforts for each player	\$
p_i	The probability of winning for the i-th player	
α	Seniority Coefficient	\$ per year
α_2	Quadratic Seniority Coefficient	\$ per year
β	Education Coefficient	\$

Some variables are not assigned values, as they are the probabilities or the constants (scores, percentage).

3.2 One Qualification Case

In the first stage, this paper is going to make a simplification about the relationship between qualifications and compensations.

The model considers the seniority as the only qualification, but it assumes that for different companies the seniority's importance varies. Quadratic relationship and linear relationship between seniority and compensation are considered depending on the data observed, and the coefficients vary for different companies.

3.2.1 Model For Company 1

For Company 1, the quadratic interpolation is used to explain the sudden increase of compensation for the employee with 25 years of seniority (Figure 1), and the equation is given as below:

$$Com = \alpha Sen + \alpha_2 Sen^2 + C \tag{1}$$

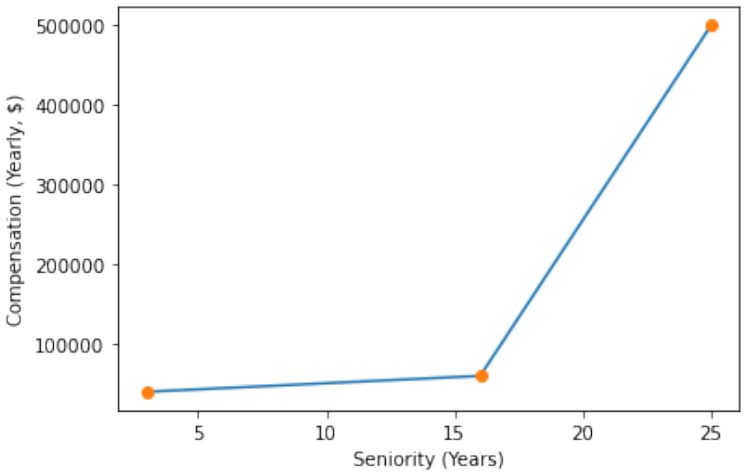

Figure 1: Company 1 data for Seniority and Compensation

Polynomial interpolation is used to get the coefficient α, α_2, and constant C, resulting in α = -39355,α_2 = 2152 and C = 138694. With the value assigned, the equation can be described as below:

$$Com = -39355Sen + 2152Sen^2 + 138694 \tag{2}$$

Figure 2 represents the relationship between seniority and compensation from the above model. By observation, the Polynomial Interpolation explains the sudden surge in salary after seniority reaches 16. However, this model doesn't perform well for seniority ranging from year 3 to year 8, since there exists a negative correlation between seniority and compensation, which is not common in the real world scenario.

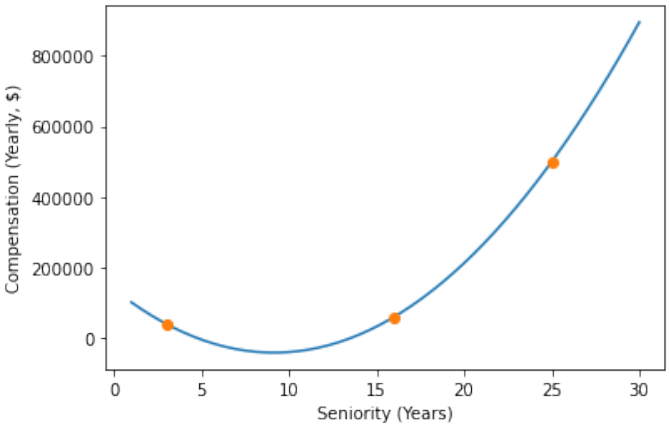

Figure 2: Polynomial Interpolation for Company 1 data

This limitation can be explained by the lack of qualifications. Since the only qualification in this section is seniority, it is reasonable that other important qualifications are not considered, leading to the discrepancy between theoretical results and real world situations.

3.2.2 Model For Company 2

For Company 2, the least square fitting is used to capture the relationship between seniority and compensation. To translate the categorical information into quantitative information, a table is provided below, and Figure 3 is presented for visualization.

The average age of NFL players is 26 years[3]. Therefore, 12 years of seniority is set to be the maximum seniority in our model, since this range is expected to cover most of the players.

Categorical	Quantitative
Rookie	year 0,1
2nd y	year 2
3rd y	year 3
4th-6th y	year 4,5,6
7th-9th y	year 7,8,9
10th + y	year 10,11,12

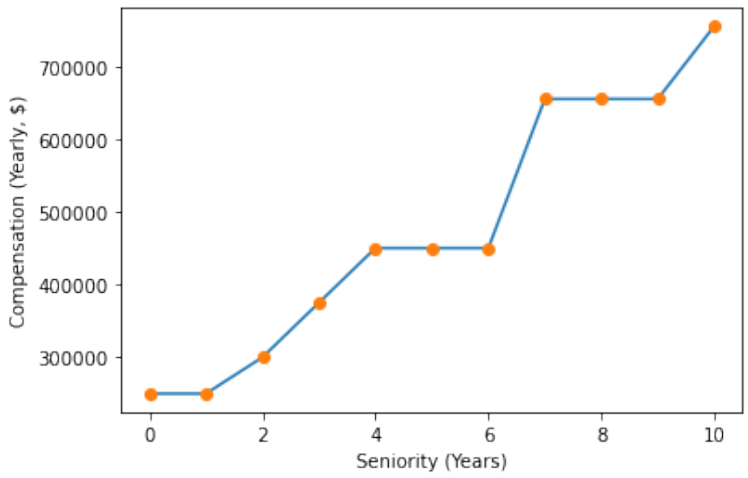

Figure 3: Company 2 data for Seniority and Compensation

The formula is given as below:

$$Com = \alpha Sen + C \tag{3}$$

As previously mentioned, least square fitting is used to derive the value for α and C, which are 52454 and 214545, respectively. With the values assigned, the formula can be described as:

$$Com = 52454Sen + 214545 \tag{4}$$

The graph below is a visual representation of this least square fitting model. By observation, the linear relationship is captured well by the model. With an R square as high as 0.94, this model fits well with the data from a statistical perspective.

Figure 4: Actual Data and Theoretical Result for Company 2

3.3 Multiple-Qualification Case

In reality, the salary is influenced by more than one factor. Thus, for the second stage, more qualifications will be added to the formula. For professors, the education is an important aspect, and it is ranked based on the Ph.D. institution of the professor. For the NFL league, the players earn salaries based on the club's revenue, and for the clubs with strong financial power, they are generous in paying more bonuses and compensation.

3.3.1 Professors

The linear regression oversimplifies this case. Although professors accumulate knowledge over time, and seniority qualification has a linear relationship with the compensation, the education which plays a key role behaves differently. Instead of a linear relationship, education, Edu, is described as an exponential decay function in this section, and gender, Gen, is considered as a dummy variable.

$$Com = (1 - Gen)(\alpha Sen + \beta e^{-(Edu-1)/r} + C) \tag{5}$$

Female professors receive 16% less than what their male counterparts do [1]. Therefore, Gen is equal to 0.16 if the professor is female and is equal to 0 if the professor is male. αSen determines the amount of compensation that

professors receive due to their seniority. Lastly, C is the base salary - the minimum salary - professors are granted.

$$\beta e^{-(Edu-1)/r} \tag{6}$$

Edu accounts for the rank of the professor's Ph.D. institution reported on the U.S. News. Function (6) is an exponential decay, and as the institution's rank gets too low, the decrease in rank will have little effect on the professor's compensation. r is the rate of compounded diminishing return, and its value indicates the sensitivity of the Ph.D. institution's rank change. β is the amount of bonus compensation the professor receives, if the professor's Ph.D. institution is ranked first, $Edu = 1$.

3.3.2 NFL Players

We set up a different model for the compensation for the NFL Players, as the club operates slightly different from corporations: members' working results, the prize, don't directly create profits for the club; instead, the clubs advertised their training efforts and past prizes to attract bonus and sponsors. As a result, we first assume that

$$S_i = r_i * s \tag{7}$$

where r_i is the share of player i on the club revenues, s is the club's revenue, and S_i is the ith player's salary. That is, the individual's salary equals his/her share of the club's total salaries timing the amount of revenues the club received from the sponsors.By considering the change of the total revenue Δs and the change of the player's share Δr, we get that

$$\Delta S = r * \Delta s + s * \Delta r \tag{8}$$

indicating the change in individual's salary is the sum of the impact of change in club's incomes and in the player's share.

In equilibrium, where the change of the share Δr is zero, we get the rate of change in individual player's salary is equal to the rate of change in the club's revenue Δs. As a result, we can calculate how individuals' qualifications impacts individuals' compensations by approaching the incomes for the club as a whole.[10] After calculating the club's income, we can easily calculate the individual's salaries by timing his/her share of the club's revenue.

As a result, the following section will first model the club's revenue s and then discuss how the club revenue helps to derive the player's compensations– that is how r, player's share of the club's revenue, works.

1. Club Revenue, π

   Basically, we introduce three variables that determine the revenues for NFL clubs: the training efforts for each player, e_i, the value of past prizes received, which we further divided into the past prize values, V_{past}, and the potential prize values, V, and finally the costs of efforts for each player, c_i. And we construct the model as :

$$\pi = R_1 * (\sum_{i=1}^{n} e_i) + R_2 * V_{past} + V * \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} c_i \tag{9}$$

   where π is the total revenue for NFL clubs, R_1 and R_2 are the revenue coefficient for the efforts and past prizes. [12] We assume that the club publicize its value to the public to receive bonus from sponsors purely via advertisements on their player's training efforts and past prizes; hence we sum up the efforts for each player and all the prizes the club has gained and time them with the correspond revenue coefficient to calculate the final revenue.

   (a) Past and future prize, V_{past} and V

   Because we are considering the club's profits as a integrity, here we include the prize for the club instead of for certain individuals. This is because most prize are earned by the corporation of all the players in the club and all of the players' qualifications decide the possibility of winning the prize. Hence, when calculating the club's revenue, we use the prize gained by the whole group timing with the sum of each member's winning possibility.

   (b) Probability of wining, p_i

   For the probability of winning for each player independently, it is decided by both the efforts that this member paid as well as the efforts that the adversaries would make. Hence, we form up the formula for p_i:

$$p_i = \frac{{e_i}^{\lambda}}{\sum_{j=1}^{N} {e_j}^{\lambda}} \tag{10}$$

where λ is the value of efforts: a relatively high value of λ represents that slightly higher effort will result in an increase in the possibility of winning the prize or the final result, while a relatively low value of λ means that efforts contribute less to the final results and other factors such as luck might play a relatively more important role. And j is the number of adversary in other clubs. By taking the division, we are looking for the weight of the results of efforts of the i-th player in our club over the efforts of the adversaries, which in turn show the possibility of winning for an individaul player.

(c) Member's effort, e_i

Specifically about the member's efforts, which are also determined by various factors. In our calculation, e_i, the effort for player i, would perfectly obey the law of individual rationality that

$$p_i * V * e_i \geq e_i, \forall i \in [1, n] \tag{11}$$

In the left side of the equation, p_i, the probability of wining a prize of value V for player i in the club, is timed with the value of prize value, V, as well as the individual's efforts, e_i. This equation informs that player will always receive at least same paybacks from the contests as the efforts he/she works during training.

As a result, our model for the club revenue will be adaptive to the players' negative qualifications such as missing goals of certain players as this negative performance will impact these players' possibilities of winning and hence impact the total revenue of clubs.

2. Player's Compensation

Now we need to derive the player's salary from the model of the club's revenue, and we need to reapply the equation 7

$$S_i = r_i * s$$

where S_i is the player's compensation, s is the club's revenue and r_i is the player's share of the club's total revenue. We here model the share of the club's total revenue based on seniority with similar approach as we applied in previous section that

$$r_i = \alpha * Sen + \beta * e_i + C \tag{12}$$

where Sen is the years of seniority and C is a constant for potential variances. In general, the share of members in the club has a linear relationship with the years of seniority and the efforts that members make in the training, deciding by the coefficient α and β respectively which varies among different clubs and situations. That means, the share of a member will increase linearly as the member becomes older and performs better in the training.

In conclusion of the individual model, the player's salary is influenced by his/her qualifications which is represented and calculated through the possibility of winning, his/her performance during the training, the prize that the club corporate to win, and the years of seniority. To be specific, better performance will directly increase the attraction rates of sponsors and hence increase the club's income and will also increase the possibility of winning a prize in the future, and negative qualifications will decrease the sponsors and the possibility of winning and hence decrease the club's income. Besides, the seniority of the member would positively impacts the member's share of the club's income and hence decide the player's salary.

4 Conclusion

4.1 Strength

The model has the following strengths:

- The two models for the compensation of academic workers and NFL players are flexible. Different institutions and teams have their own priorities; therefore, the models allow them to adjust the model easily by replacing the values of the coefficients.

- Each of the models is built for a specific field (academic workers or NFL players). This happens due to the differences between the two fields, so the models predict the salary of the people more precisely than what other generalized models do.

- The models are practical. The formulas written in the report closely resemble the institutions' or the clubs' decisions in reality. Thus, they are useful for forecasting one's true value.

4.2 Weakness

- The model for company 1 in stage 1 encounters negative correlation between seniority and compensation, which is a result of the oversimplified assumption that seniority is the only qualification.
- Due to difficulty of data collection, the paper doesn't provide detailed analysis for stage 2 models(academic workers and NFL players).
- Some variables are hard to quantify, which might require further adjustment (e.g. potential prize values V)

4.3 Future Work

In general, our model requires more supports from real-world data. In the future, the model needs to be tested with appropriate data sets on the salaries and the qualifications. With these data, we can perform precise statistical calculation like regression to find out the exact values for our coefficients and their trends for different occupations and various level of seniority.Besides, sensitivity analysis is also needed. The model should be tested in multiple situations to increase the complexity of our model. For instance, more variables or more detailed correlations need to be added when we try to apply our model with bigger database or apply our model to analyze the correlation between seniority and compensations for IT workers.

As a result, our model can be more sensitive and authentic in predicting the relationship between each qualification and the compensation if more data is included and tested.

References

[1] Alexia Aston. Ou professor finds gender wage disparities in research as provost's office evaluates equity in salary data, Apr 2021.

[2] Marcia L Bellas. Faculty salaries: Still a cost of being female? *Social Science Quarterly*, 74(1):62–75, 1993.

[3] Leslie Bloom. MS Windows NT how long is the average career of an nfl player?, 2019.

[4] Sayan Das, Rupashri Barik, and Ayush Mukherjee. Salary prediction using regression techniques. *Proceedings of Industry Interactive Innovations in Science, Engineering & Technology (I3SET2K19)*, 2020.

[5] Frank M Howell and Deborah Richey Bronson. The journey to work and gender inequality in earnings: A cross-validation study for the united states. *Sociological Quarterly*, 37(3):429–447, 1996.

[6] Ita GG Kreft and Jan De Leeuw. The gender gap in earnings: A two-way nested multiple regression analysis with random effects. *Sociological Methods & Research*, 22(3):319–341, 1994.

[7] Nancy Langton and Jeffrey Pfeffer. Paying the professor: Sources of salary variation in academic labor markets. *American sociological review*, pages 236–256, 1994.

[8] David J Maume Jr. Wage discrimination over the life course: A comparison of explanations. *Social Problems*, 51(4):505–527, 2004.

[9] M Navyashree, MK Navyashree, M Neetu, GR Pooja, and Biradar Arun. Salary prediction in it job market. *Int. Journal of Comp. Sci. and Engineering*, 7(15), 2019.

[10] Gerald W Scully. Player salary share and the distribution of player earnings. *Managerial and Decision Economics*, 25(2):77–86, 2004.

[11] Ryan A Smith. Race, gender, and authority in the workplace: Theory and research. *Annual Review of Sociology*, 28(1):509–542, 2002.

[12] Stefan Szymanski. The economic design of sporting contests. *Journal of economic literature*, 41(4):1137–1187, 2003.

[13] Kim A Weeden. Why do some occupations pay more than others? social closure and earnings inequality in the united states. *American journal of sociology*, 108(1):55–101, 2002.