# Player Performance Prediction

## Theory and Method

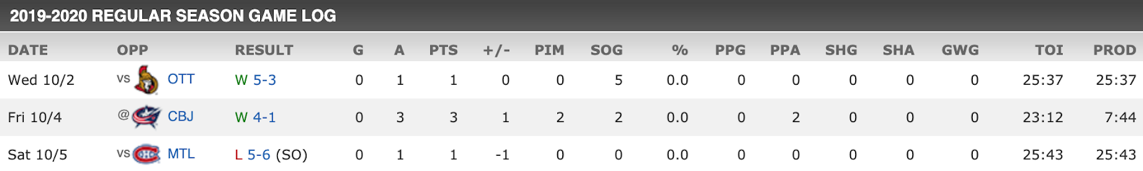
In the national hockey league, every team is like a individual business, and the salaries paid to players would be a big part of business cost. As for Boston Bruins and Toronto Maple Leaf, their hold different philosophies about how to hire players. Boston Bruins pays for player according to their performance in the games, but Toronto Maple Leaf is famous for its paying strategy that it pays to the players who have very good prospects in the future a lot. This discrepancy shows in the difference between the salary structures of these two team.

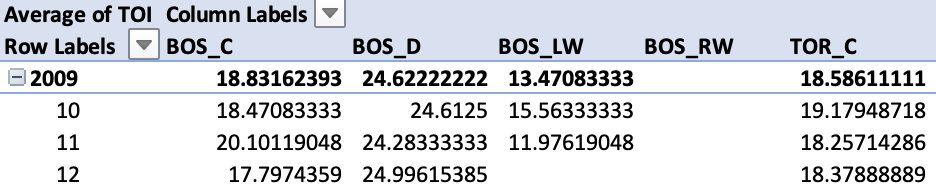
The top-five players in Boston just occupy $30 million of the total salary, but the top-five players in Toronto occupy $45 million of the total salary. The total salary of each team is limited by the league, which means that if the team is willing to pay more for their star players, it has to pay less for the rest. It is a kind of trade-off. We would like to see if Boston Bruins pays a fair amount to the players, or the players in Toronto Maple Leaf worth the salaries in the future. The matrix we selected to evaluate players is Time On Ice (TOI), which measures total amount of time a player is on the ice during the playing of the game. It is an indicator of how a coach views the player and their contribution with more time going to players considered to be producers either through point production or point prevention.

The players we will describe are the top-one player at each position in each team. So there would be eight player in this analysis, they are the centerman, defenseman, left wing and right wing of each team. Due to space limitations, we would just make prediction on the centerman and defenseman.

## Data Processing

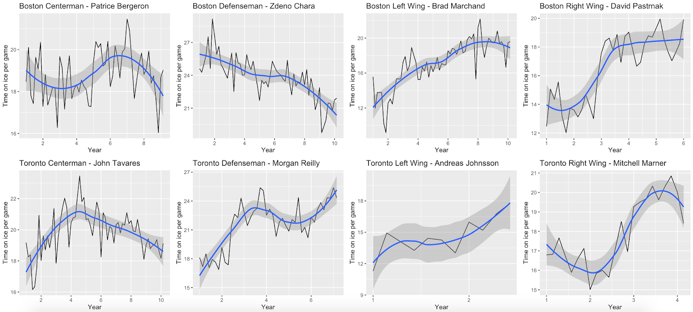
1. We extracted 10-year game log data of each player from <https://www.nhl.com> and calculate the monthly average of Time On Ice (TOI) in Excel by using pivot table.



1. The table shows that some records are missing. To solve this problem, we filled out them by using average TOI of surrounding four months. Besides, we have to remove the records of season 2012-2013 with just four months due to due to a lockout imposed by the NHL franchise owners.
2. After we cleaned the data, the performance of last year (7 observations) would be used as test set.

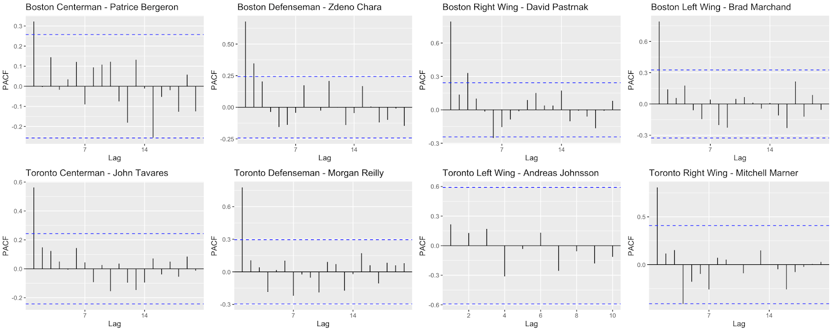
## Data Description

1. So now we get the date ready for analyzing. There are 7 months in each season, so there will be around 70 observations for each time series. Let’s see how they looks like:



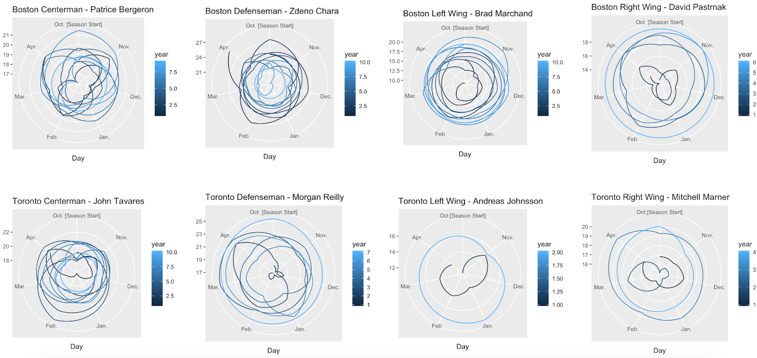
The trend and range of each player’s TOI is different. Typically defenseman will generally log the most amount of time on ice usually around 20+ min while the top forwards on the team will log around 17-21 min per game. And the old player, like the Boston Defenseman, Zdeno Chara, plays less and less in the last ten years. But the young player, like the Toronto Defenseman, Morgan Reilly, plays more and more since he get into the national hockey league.

1. PACF plots show us the relationship between the performance in neighboring months. Since it is obvious that the significance lags of most of the graphs are lag 1, except for the one for player Andreas who is a new player without much data to show trends, we could see a strong relationship between the performance in neighboring months.



1. Seasonality is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly. Polar maps of the performance of 8 players in Boston Bruins and Toronto Maples show us which month the player has better performance and seasonal performance stationarity.

The difference between line to the middle of the polar map is the level of TOI. Some players like Brad Marchand has a stable performance through every season. But the player, John Tavares always plays better at the middle of the season than that the beginning. On the other hand, Patrice Bergeron plays well at the beginning of the season but goes through a trough at the middle season.



## Player Performance Forecasting

### Tests of Non-linearity

We checked the linearity of each time series to see if the linear model is applicable.

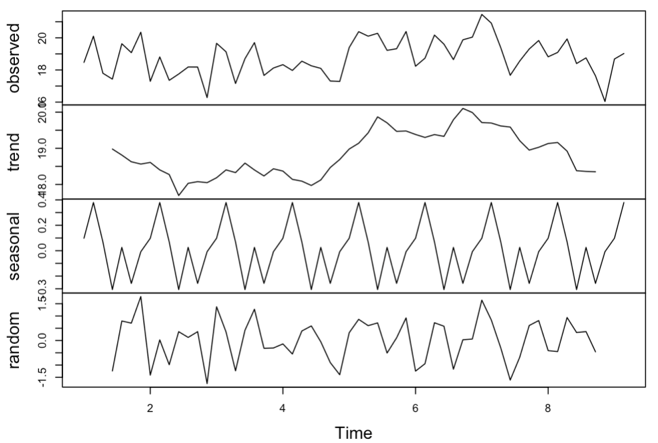
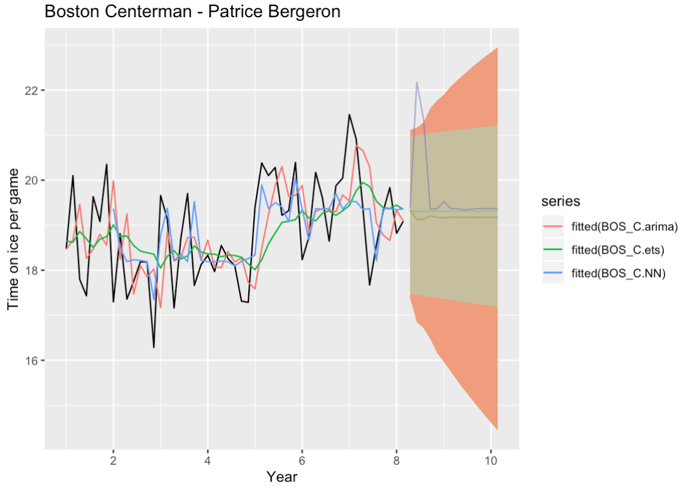
* Boston Centerman - Patrice Bergeron passed 5/6 tests
* Boston Defenseman - Zdeno Chara passed 4/6 tests
* Toronto Centerman - John Tavares passed 4/6 tests
* Toronto Defenseman - Morgan Reilly passed 3/6 tests

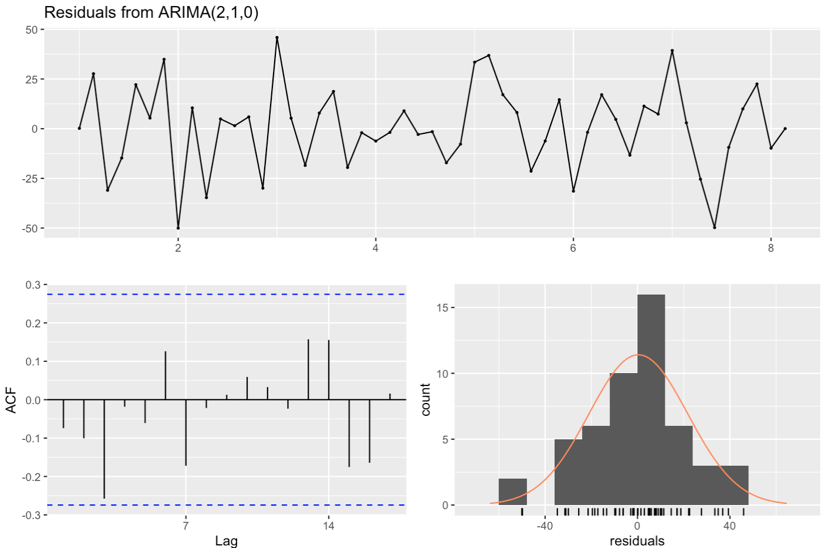
Linear model is not so suitable for the analysis of Morgan Reilly’s performance.

### Boston Centerman - Patrice Bergeron

Patrice is the alternative capitan of Boston Bruins, an all-star player. His biggest advantage is that he can play well in both defensive side and offensive side, so he is a kind of transition player, which is very important for the team.

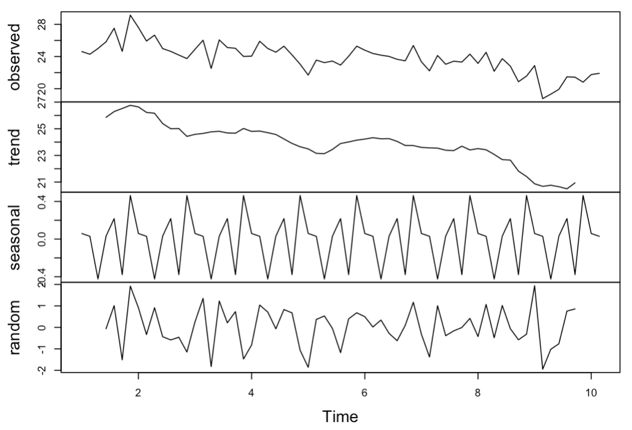
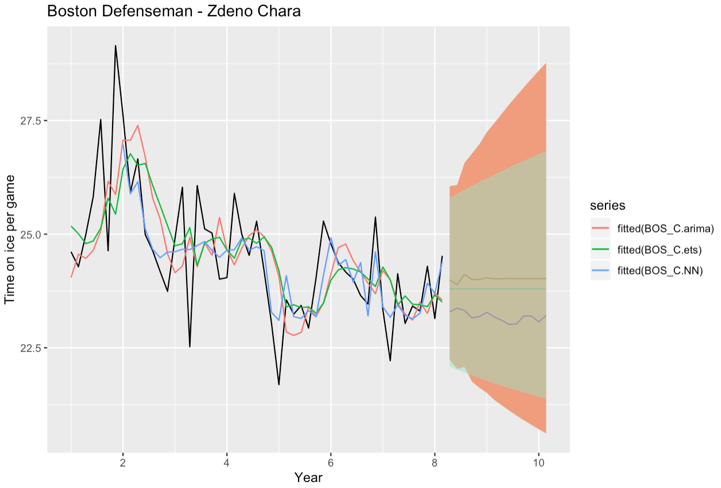
He is aged 34 years old, just passing the golden age of hockey player and he get $8.75 million each year. We can see an overall performance trend in the past 10 season from decomposed graph.



 According to the MAPE and MASE of test set, we choose ARIMA(2,1,0) to predict his performance in the next two year. The model gives a constant forecast, which suggest the team manager to pay the same amount to Patrice.

### Boston Defenseman – Zdeno Chara

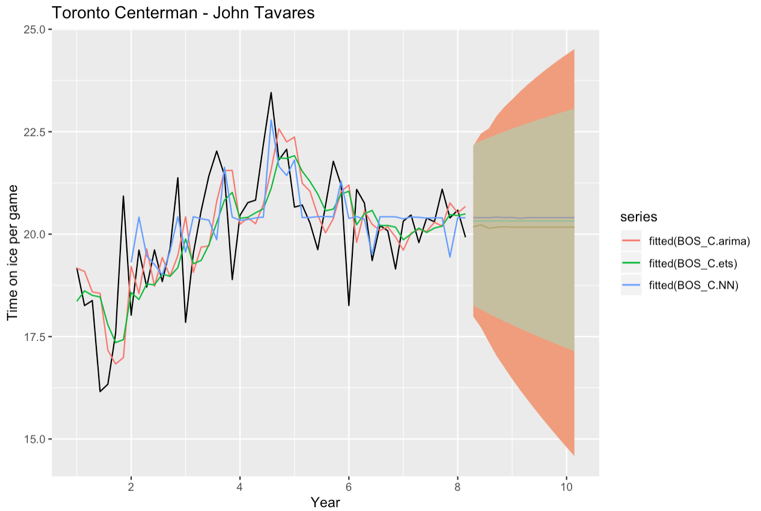
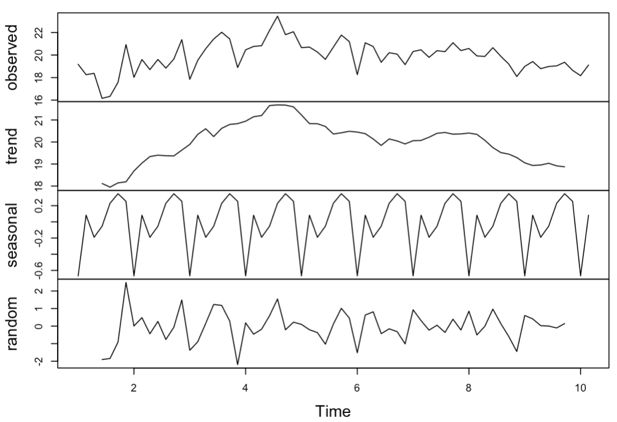
Zdeno is the Capitan of Boston Bruins, playing in national league for 24 year. He is addicted to exercise, so even he is 42 years old now, he still keeps a good shape. When he was young, his goal is playing all over each game, 60 mins. Even it is no possible, he is still capable, playing around 20 minutes per game.



As we can see the trend curve from the decomposition, there is an obvious downward trend in his performance. And according to the MAPE and MASE of each model on the test set, we would like to use neural network model to predict his performance. As expected, Zdeno will play less and less in the next two season, and we would suggest the team manager to reduce his salary.

### Toronto Centerman - John Tavares

John is a star player since he got into the league. He ranked 1st in the draft, indicating huge potential in future, which is one of the reasons why his salary could be so high, $11 million per year. As we mentioned before, Toronto pays for its player based on their future prospects. This huge amount of salary indicates that that would be a huge improvement with John. Unlike the centerman in Boston Bruins, John focuses on offensive and his competitive state get stable and stable in the past two season.



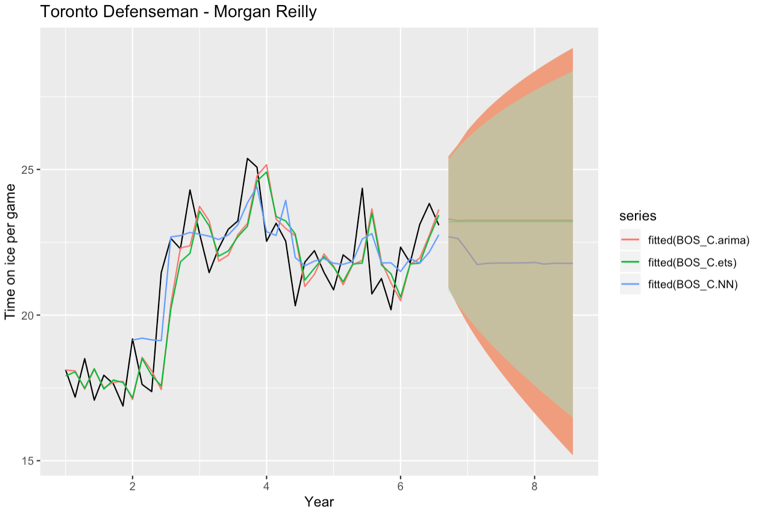
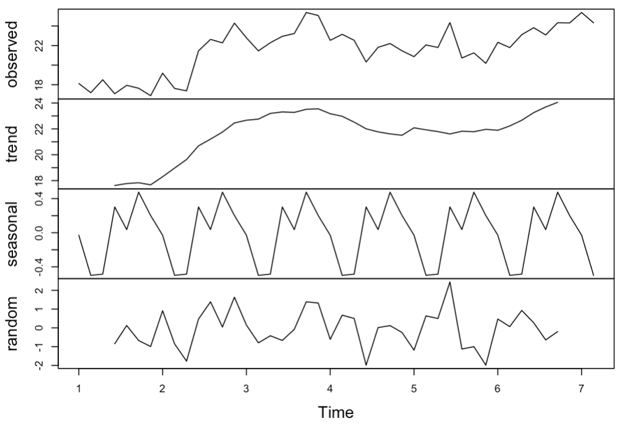
A picture containing text

Description automatically generated By comparing the MASE and MAPE, we decide to use ARIMA(2,1,0) to predict John’s performance in the next two season. It gives us a constant prediction, which contradict with team manager’s expectation of John. So we would suggest the team manager to slightly reduce the salary of John.

The residual plot looks really nice, no auto-correlation issue exists.

### Toronto Defenseman - Morgan Reilly

Morgen is the top defenseman in Maple leaf and he is very young. We can see there is a significant improvement within his performance six years ago. In the ice hockey game, defensemen usually play better and better as they gained more and more experience, being familiar with how opponents play the game.



According the MAPE and MASE, neural network can make a better prediction. It implies that the Morgan’s performance is likely to go down in the rest of this season. Since his salary is not so high, and his is very young, we would recommend the team manager to keep the same salary in the next two years.

### 