Team Talent Imbalance in the NHL

SFU | Fall 2018 | STAT 485 | Steven Thompson

Trevor Dallow

Contents

[Introduction 3](#_Toc531565129)

[Methods 4](#_Toc531565130)

[Normality 4](#_Toc531565131)

[Stationarity 4](#_Toc531565132)

[Model Specification 4](#_Toc531565133)

[Parameter Estimation 5](#_Toc531565134)

[Model Diagnostics 5](#_Toc531565135)

[Results 6](#_Toc531565136)

[Discussion 6](#_Toc531565137)

[References 7](#_Toc531565138)

[Appendix 8](#_Toc531565139)

Table of Figures

[Figure 1: Q-Q plot of original data. 8](#_Toc531565151)

[Figure 2: Q-Q plot of log transform. 8](#_Toc531565152)

[Figure 3: Plots of original and 1st differencing 9](#_Toc531565153)

[Figure 4: ACF of log transform of data. 9](#_Toc531565154)

[Figure 5: PACF of log transform of data. 10](#_Toc531565155)

[Figure 6: Residuals of simulated model. 10](#_Toc531565156)

[Figure 7: Q-Q plot of residuals of simulated data. 11](#_Toc531565157)

[Figure 8: ACF of residuals. 11](#_Toc531565158)

[Figure 9: Plot of original and simulated time series. 12](#_Toc531565159)

# 

# Introduction

The National Hockey League (NHL) recently began another cycle of expansion, first with the addition of the Los Vegas Knights to the league in the 2017/18 season and next with a team based in Seattle which will join in the 2020/21. Historically, teams that have been newly introduced to the league have been plague with years of poor performance, having been dominated by the more tenured teams in the league. Recently however, it seems that the NHL has determined a method of introduction that spares new teams from years of growing pains, as was seen with the Los Vegas Knights successful campaign in the 2017/18 season. In this report we investigate how the imbalance of team talent across the league evolves over time in terms of the difference in points acquired by the best and worst performing teams for each season.

The raw data was acquired from the Internet Hockey Database where 100 seasons of NHL standings data spanning from 1918 to 2018 was scrapped from the website [HockeyDB]. The standings for each season include data such as team name, number of games played, number of wins, number of losses, number of points, etc. for all the teams the competed in that season. The data of interest for this report is the number of games played and number of points won by each team for each season. Over time, the number of teams participating in a given season and the number of games played for each season fluctuates, so the time series data of interest must ensure that each data point must be on the same scale. In the next section we will discuss the time series data specifically.

The time series data in this report uses the difference in points won by the best and worst team to represent a metric for the league imbalance of each season. For each team in a given season, the proportion of points won out compared to the total number of points that could possibly be won is computed. Then the difference in these proportions of the best and the worst performing team is taken for each season. After which the data is normalised such that the scale is the difference in the number of points won between the best and worst team as if the teams played 82 games in a season. 82 games are used as it is currently the standard number of games played by each team per season. Note that there is no 2005 season due to a lockout resulting in a cancelled season.

# Methods

## Normality

When checking for normality in out time series data, we see significant evidence against it given the curvature of thee QQ-plot [Figure 1]. Doing a log transform improves the normality of the data significantly [Figure 2]. We will proceed with the log transform of the data for further analysis.

## Stationarity

Now we wish to specify a model for this time series data. First we examine the stationarity of the time series data. In this instance the data appears to be stationary without the need for any difference [Figure 3]. We also consider the results of an Augmented Dickey-Fuller Test which checks for the existence of unit roots in the characteristic polynomial. With an alternative hypothesis that the time series is stationary and testing with a lag of order 0 we get a p-value of 0.01, which with an = 0.05 gives us significant evidence to reject the null hypothesis that the process is non-stationary and conclude that the process is stationary without the need for differencing. This result supports our findings from the time series plots.

## Model Specification

Next we wish to consider the ACF and PACF with regards to our time series data to determine the p and q parameters for our ARIMA(p, d, q) model. Examining the ACF plot we that the sample acf values seem to trail off after lag 1 which may suggest an MA(1) model [Figure 4]. The acf values at lag 10 exceed the error bound, this will be ignored for a more parimonious model. As well, the PACF values of the PACF plot remain within the error bounds after lag 1, except for lag 10 [Figure 5]. For the purposes of parimony we will only consider up to an AR(1) model.

We also conduct hypothesis tests to determine the order of the model.

: is a MA(q) model given for k > q

: is a AR(p) model given for k > p

We see from the results of our testing that we do not reject MA(1) and AR(1) models, supporting our findings from analysing the acf and pacf plots.

## Parameter Estimation

Now that we have a general idea of the order of our model, we may start to estimate the parameters of said model. We consider two methods for estimating the parameters: Least Squares and Maximum Likelihood Estimates. In addition to ARMA(1, 1), we also consider AR(1) and MA(1) models. The results from our parameter estimation tests agree with our suspicions that the data is an ARMA(1, 1) process given that the log-likelihood for both methods is maximized when an ARMA(1, 1) model is considered.

## Model Diagnostics

We start our model diagnostics by checking whether the residuals express any trends. We see that our residuals are centrered on a zero horizontal line and do not appear to have any trend [Figure 6].

Next, we check the normality of the residuals using Q-Q plots [Figure 7]. The residuals appear to roughly follow a normal distribution, though they may be heavy-tailed. There is a significant outlier on the left tail of the distribution. By doing a Shapiro-Wilk Normaility Test we can statistically evaluate the normality of the residuals. With a p-value = 0.077, we do not reject the normality of the residuals ().

: The sample data comes from a normal distribution.

We wish to evaluate the autocorrelation of the residuals. Given the ACF plot of the residuals we do see that there is some correlation between them, specifically for lags 10 and 12 as they exceed the boundaries [Figure 8]. This is cause for concern as it suggests there exists some autocorrelation between residuals and so the error terms may not be entirely white noise.

We conduct a Ljung-Box Test to check whether the error terms are uncorrelated. At we fail to reject the hypothesis that the data are independently distributed (p-value = 0.579). WIth this result we may consider that the error terms are uncorrelated.

: The data are independently distributed, i.e. error terms are uncorrelated.

# Results

The final model chosen to fit the time series process is an ARMA(1, 1) model with parameters and (following the textbook’s convention of MA models) [Figure 9].

# Discussion

Given that parameter estimates suggest a single AR component, this indicates that the point differential of the previous season is slightly influences the current season’s point differential. This can be explained by the fact that teams typically don’t change much as far as which players are on the team and coaching styles from one season to the next. As a result one may expect the results of the previous season to yield some insight as to how the team will perform in the current season. But over the course of 5 years for example, many players may be traded, retire or brought up from farm teams as well as changes in the coaching staff which result in the same team being significantly different when compared 5 years apart. It is not typical for the composition of a team to change significantly in the span of a few seasons so we would expect a low order for the AR component of the ARMA model.

Future improvements could be to decide on a different metric for comparing the performance of strong and weak teams. In this report, only the best and worst team for each season were compared which may not be indicative of league imbalance in that one team could simply have an unusually good or bad season but the rest of the league was relatively competitive. One could instead consider a metric that compares the points obtained by the best set of teams against the worst set of teams to mitigate the impact of outliers. As well, instead of just looking at points acquired, other team statistics could be investigated such as goals scored for and against and compare the best and worst teams.

# References

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis With Applications in R.* New York: Springer.

HockeyDB. (n.d.). *http://www.hockeydb.com/ihdb/stats/leagues/141.html.* Accessed 2018-11-26.

National Hockey League (1 December 2018). In Wikipedia. Retrieved 26 November, 2018, from https://en.wikipedia.org/wiki/National\_Hockey\_League.

# Appendix

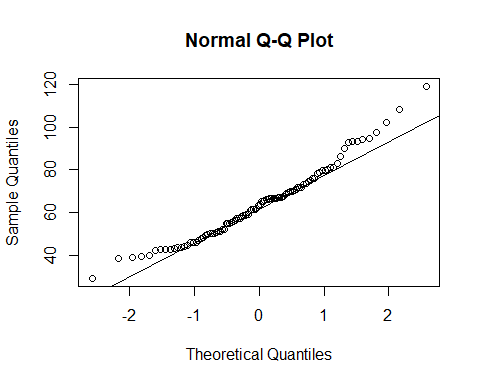


Figure 1: Q-Q plot of original data.

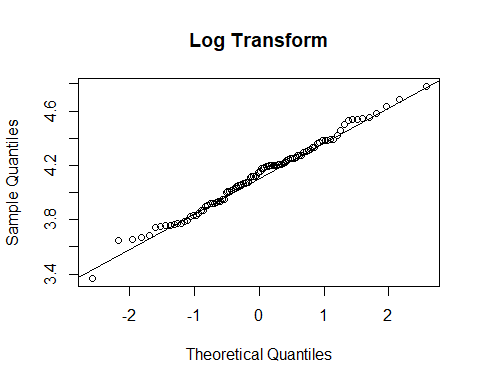


Figure 2: Q-Q plot of log transform.

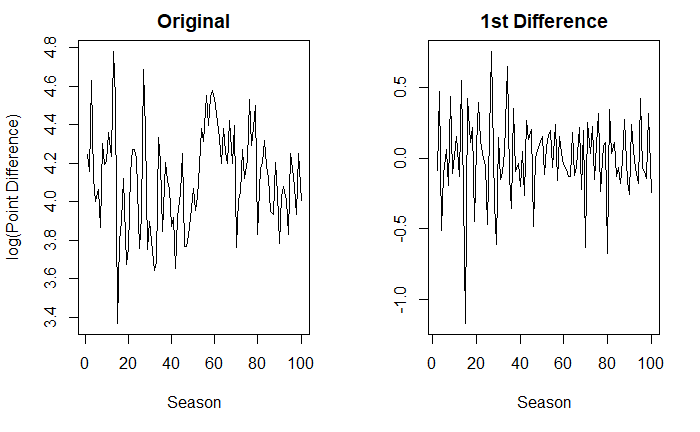


Figure 3: Plots of original and 1st differencing

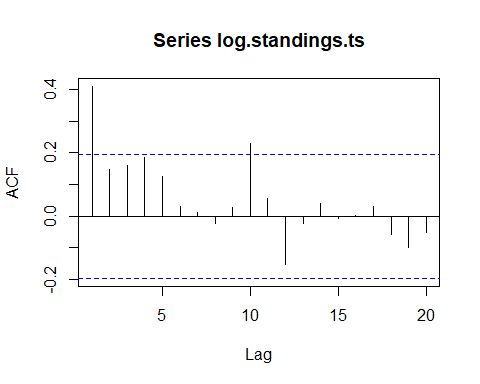


Figure 4: ACF of log transform of data.

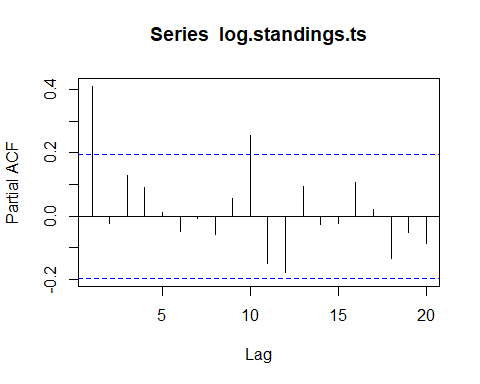


Figure 5: PACF of log transform of data.

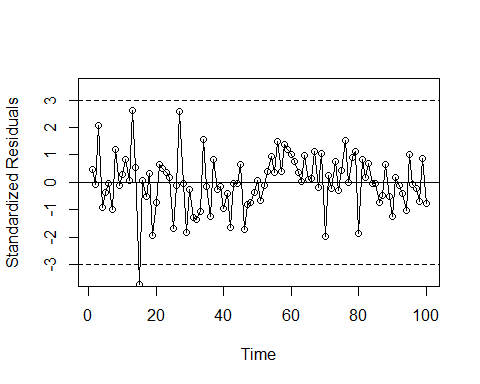


Figure 6: Residuals of simulated model.

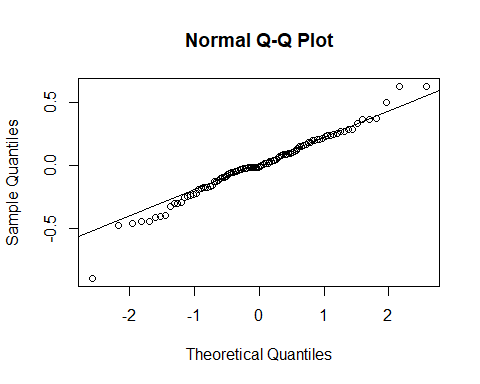


Figure 7: Q-Q plot of residuals of simulated data.

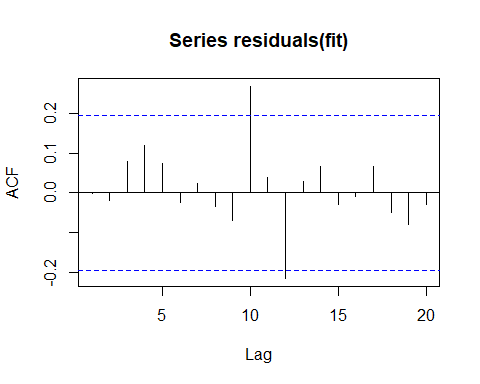


Figure 8: ACF of residuals.

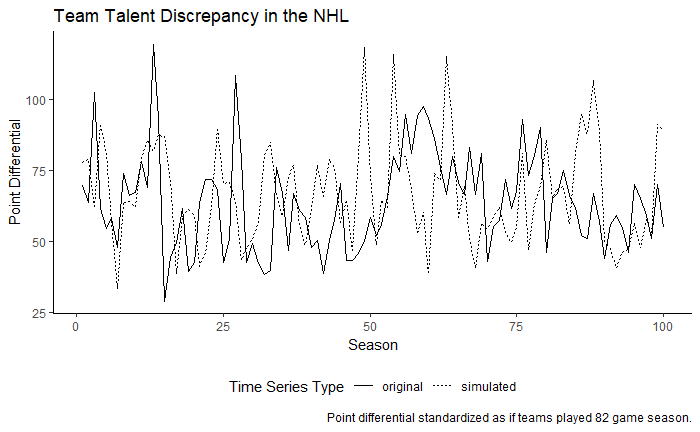


Figure 9: Plot of original and simulated time series.