Multiple Regression Analysis: Understanding the Impact of Offensive and Defensive Contributions to Team Performance

By Kevin D. Dayaratna and Steven J. Miller

From evaluating team performance to predicting outcomes of particular matchups to understanding individual player contributions, rigorous statistical analysis has become increasingly useful in hockey over the past several decades. In recent years, there has been a variety of research projects offering quantitative assessments from a variety of perspectives. This study adds to those efforts.

A number of recent studies have provided quantitative methods for evaluating team performance. For example, in 2013, Kevin Dayaratna and Steven J. Miller looked at the use of the well-known Pythagorean Won-Loss formula for evaluating team performance (Dayaratna and Miller 2013). They showed that the model, initially intended for baseball, was just as appropriate for hockey. Subsequently, Kevin Dayaratna and Max Mulitz bridged the 2013 Dayaratna and Miller study with another study by Dayaratna and Miller in 2012, showing that a well-known linear model is just a first order approximation to the Pythagorean Won-Loss formula (Jones and Tappin 2005; Dayaratna and Miller 2012; Dayaratna and Mulitz 2014). They showed that this linear model, also initially intended for baseball, works just as well for hockey. McDonald (2014) extended the Dayaratna and Miller (2012) study by estimating a similar linear model to baseball, but used runs scored and runs allowed as separate explanatory variables, enabling them to understand the impact of runs scored and runs allowed on winning percentage. In this study, we apply the linear model from the McDonald (2014) study to NHL data to understand the impact of goals scored and goals allowed on overall outcomes.

Model Development

In this section, we apply the linear model from the McDonald (2014) study to baseball. The linear model presented in McDonald (2014) is:

$$WP \approx \alpha + \beta_1 GS + \beta_2 GA$$

The response co-efficients â, $\hat{\beta}_1$,and $\hat{\beta}_2$, which can be estimated via the method of least squares, indicate the impact of the explanatory variables on the dependent variable. In particular, $\hat{\beta}_1$ and $\hat{\beta}_2$ indicate the impact of offence and defence on performance. The value of $\hat{\beta}_1$ describes the marginal change of the winning percentage with respect to how many goals a team scores. A value of 0.10, for example, would suggest that a team's winning percentage increases by 10% when they improve their average number of scores by a single goal. Similarly, the value of $\hat{\beta}_2$ describes the marginal change in winning percentage with respect to goals allowed.

Data/Results

Our data was downloaded from ESPN.com for each NHL team over the 2005/06, 2006/07, 2007/08, 2008/09, 2009/10, and 2010/11 seasons (ESPN.com). We estimated the multiple linear regression model discussed in the previous section in R. Table 1 contains our coefficient estimates with p-values in parentheses: (see over)

Table 1: Co-efficient Estimates and Model Fit Statistics

Season	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2
2005/06	0.47949 (<0.01)	0.17744 (<0.01)	-0.17128 (<0.01)	0.8338 (<0.01)
2006/07	0.68221 (<0.01)	0.13095 (<0.01)	-0.19330 (<0.01)	0.8691 (<0.01)
2007/08	0.57605 (<0.01)	0.15618 (<0.01)	-0.18259 (<0.01)	0.8753 (<0.01)
2008/09	0.49870 (<0.01)	0.18469 (<0.01)	-0.18382 (<0.01)	0.8825 (<0.01)
2009/10	0.57217 (<0.01)	0.17426 (<0.01)	-0.19944 (<0.01)	0.8909 (<0.01)
2010/11	0.54108 (<0.01)	0.16469 (<0.01)	-0.17930 (<0.01)	0.8554 (<0.01)

All of the co-efficient estimates above were highly significant (having p-values well below a critical threshold of 0.01). Our results are quite interesting. Most notably, in four of the six seasons (2006/07, 2007/08, 2009/10, and 2010/11), the marginal change of winning percentage with respect to goals allowed, $\hat{\beta}_2$, is greater in absolute value than the marginal change of the winning percentage with respect to the goals scored, $\hat{\beta}_1$. These results indicate that, for those seasons, teams' winning percentages were more sensitive to goals allowed than goals scored, signifying that defence played a more significant role in the NHL than offence.

There are a variety of implications this type of regression analysis can have. For example, hockey teams, at all levels, can use this type of analysis to understand reasons behind how they are performing. Additionally, teams can use this type of analysis strategically against other teams, to determine whether offence or defence is more important in determining how their opponents perform. For example, if a team determines that its opponent is likely winning most of its games as a result of its defence, they can then determine appropriate offensive strategies to challenge this defence. Similarly, if a team determines that its opponent wins most of its games as a result of a strong offence, it can fortify their defense appropriately.

Future Research

As with any study, there are a number of ideas to extend this research. Firstly, we made the assumption that all 82 hockey games across the NHL season were realizations of independent and identically distributed random variables. In practice, one may be interested in weakening this assumption, perhaps by applying a Bayesian perspective to this problem, and understanding the consequent impact on coefficient estimates. Additionally, this model treats games that ended in overtime in the same manner as games that ended in regulation time. A potential avenue of future research could be to classify overtime games as a third outcome – perhaps a tie. It could be interesting to see the impact on this reclassification on the coefficient estimates.

Finally, there is no reason that this model needs to be restricted to hockey as there are myriads of other sports that could benefit from logistic regression analysis including baseball, football, basketball, and even tennis. It would be not only fascinating, but also useful, to apply a similar model to these settings.

References

Dayaratna, K and Miller, S.J, 2012. "First Order Approximations of the Pythagorean Won-Loss Formula for Predicting MLB Teams' Winning Percentages," By the Numbers: The Newsletter of the Society of American Baseball Research and an expanded version with an alternate derivation is available at http://arxiv.org/abs/1205.4750.pdf.

Dayaratna, K and Miller, S.J, 2013. "The Pythagorean Won-Loss Formula and Hockey: A Statistical Justification for Using the Classic Baseball Formula as an Evaluative Tool in Hockey," The Hockey Research Journal A Publication of the Society for International Hockey Research. An expanded version is available at http://arxiv.org/abs/1208.1725.pdf.

Dayaratna, K and Mulitz, M, 2014. “Linear Predictors of NHL Team Performance: A Simple Statistical Model for Predicting NHL Team’s Winning Percentages” The Hockey Research Journal A Publication of the Society for International Hockey Research.

ESPN. “NHL Expanded Standings — 2008/09.” ESPN: The Worldwide Leader in Sports. http://espn.go.com/nhl/standings/_/type/expanded/year/2006

ESPN. “NHL Expanded Standings — 2009/10.” ESPN: The Worldwide Leader in Sports. http://espn.go.com/nhl/standings/_/type/expanded/year/2007

ESPN. “NHL Expanded Standings — 2010/11.” ESPN: The Worldwide Leader in Sports. http://espn.go.com/nhl/standings/_/type/expanded/year/2008

ESPN. “NHL Expanded Standings — 2008/09.” ESPN: The Worldwide Leader in Sports. http://espn.go.com/nhl/standings/_/type/expanded/year/2009.

ESPN. “NHL Expanded Standings — 2009/10.” ESPN: The Worldwide Leader in Sports. http://espn.go.com/nhl/standings/_/type/expanded/year/2010.

ESPN. “NHL Expanded Standings — 2010/11.” ESPN: The Worldwide Leader in Sports. http://espn.go.com/nhl/standings/_/type/expanded/year/2011.

Jones, M. A. and Tappin, L. A, “The Pythagorean Theorem of Baseball and Alternative Models,” The UMAP Journal 26.2 (2005), 12 pages.

McDonald, John (2014). “Extensions of the Linear Runs-To-Wins Model” By the Numbers: The Newsletter of the Society of American Baseball Research

Miller, Steven J., Eric T. Bradlow, and Kevin Dayaratna. “Closed-form Bayesian inferences for the logit model via polynomial expansions.” Quantitative Marketing and Economics 4.2 (2006): 173-206.

Miller, S.J., “A Derivation of the Pythagorean Won-Loss Formula in Baseball,” Chance Magazine 20 (2007), no. 1, 40--48. An abridged version appeared in The Newsletter of the SABR Statistical Analysis Committee 16 (February 2006), no. 1, 17--22, and an expanded version is available at http://arxiv.org/abs/math/0509698.pdf