Weighted Analytics – What Do the Numbers Suggest?

Craig Peterson, Vinodh K. Chellamuthu, & Joseph Lovell

Dixie State University

J E S S

Critical Commentary

Journal of Emerging Sport Studies

Volume 3, (2020): London, Ontario

 ISSN – 2562-3184

Introduction

The growing popularity, as well as the profit-oriented business of sports, has recently led to an increased need to understand player and team performance. The expansion of statistical analysis for predicting individual and team success continues to be a need for sport organizations (Bedford & Baglin, 2009; Chan, Chow, & Novati, 2012). This has led to the introduction of various metrics that help teams be more successful by properly evaluating player performance. As analytical methods evolve because of the larger role they play in the success of sport performance, there is a strong push to improve the methods used for evaluating individual and team performance (Moskowitz & Wertheim, 2012). Hockey and basketball, particularly, have seen a surge of using analytics as a crucial component in designing winning strategies. Using various analytics provides coaches with better insight for practices and game preparations, as they highlight both the strengths and weaknesses of teams. This in turn gives fans better insight into potential game outcomes, thus increasing the popularity and profit of the sport. This critical commentary will discuss the use of analytics in sport and provide two possible models in the advancement of analytics in both basketball (modified effective field goal percentage) and hockey (modified Corsi) for effective evaluation of individual and team performance.

Statistics is a powerful tool used to understand and analyze dynamics in the sports world. Statistical analysis assists sport industry professionals in decision making to improve teams' performance and recruitment. However, when it comes to statistics, there is an overabundance of information. Due to the vast amount of raw data available, it is easy to get lost. In order to understand statistical methods in sports, one must understand the basics of the Pythagorean Theorem, sometimes referred to as the Pythagorean expectation or Pythagorean value. While many might recall the application of this theorem for triangles, the statistic has been applied to various situations in sport. In the nineteen-eighties, Bill James developed a Pythagorean formula for baseball, focusing on the analysis of runs scored and runs allowed, as depicted in the formula: $\frac{Runs\ scored^2}{Runs\ Scored^2 + Runs\ Allowed^2}$. This formula will essentially indicate a team's level of performance: the higher the percentage, the better the team is performing. Understanding the basic principles of this formula allows team managers and coaches to make changes in their lineup throughout the season. By doing so, management can increase their team's Pythagorean value and increase the potential of

success on the field, which is ultimately the goal. This statistic provides a relatively accessible understanding of sport analytics, for both executives and casual fans.

There are a number of ways to increase a team's Pythagorean value, but the basic idea is to increase the runs scored. By increasing the number of runs scored, the Pythagorean value increases. Thus, the winning percentage increases. Another option is to lower their runs allowed. Ideally, in an optimal situation, a team would increase the runs scored while decreasing their runs allowed. Though none of this information is new, it is important to understand this basic concept before expanding on the role of statistics in sports. Additionally, several studies have shown that the Pythagorean estimate also works in basketball, football, soccer, and hockey (Cochran & Blackstock, 2009; Horowitz, 1994; Hamilton, 2011; Kubatko et al., 2007).

Basketball Analytics

Basketball has seen a significant increase in the use of analytics, which include offensive and defensive ratings, true shooting percentage, and individual efficiency, as well as many other analytical models. In particular, the Effective Field Goal (EFG) percentage is a model of great interest. The EFG is one of the major factors in Dean Oliver's Four Factor Offense (Kubatko et al., 2007). EFG is calculated as $\frac{Field\ Goals\ Made + .5(3\ Point\ Field\ Goals\ Made)}{All\ Field\ Goals\ Attempted}$ This statistic is crucial, since it gives both teams and individuals' weight, based on how they are performing (shooting). Weight is given to the three-point field goal, as it is worth fifty percent more than the two-point field goal. However, the issue with this formula is that there is not an appropriate weight given to the individuals shooting the ball, since all players have their own unique attributes, behaviors, and effectiveness when playing the game. Hence, the authors designed a novel Weighted Effective Field Goal (WEFG) percentage formula that accounts for individuals' attribute (shooting percentage) compared to EFG: WEFG =

$$\frac{Field\ Goal\ Percentage\ (Field\ Goal\ Attempted) + 3\ Point\ Field\ Goal\ Percentage\ (3\ Point\ Field\ Goal\ Attempted)}{All\ Field\ Goals\ Attempted}$$

If a team were to have the majority of their shots taken from outside the arc (three-point shot), the team would need to ensure that they have a good three-point shooting percentage on both the individual level and the team average. If the team has a low three-point shooting percentage, having a coaching strategy to shoot more frequently from the outside would be counterproductive. For example, if a team were to have a slightly below average shooting percentage from inside the arc and an average three-point shooting percentage, the optimal strategy would be to work with the coaching staff to find a system where the team would be taking slightly more shots from the outside, rather than the inside. Furthermore, if a teams' opponents have a low three-point percentage, the idea would be to lower the outside pressure and, in return, focus on the pass or rebound. Conversely, if a team had an excellent three-point percentage and an average two-point percentage, the focus would be to increase the outside pressure while minimizing the potential for passing options inside the arc.

The beauty of the WEFG is that it shows how successful the team can be, depending on the design of the team's strategy. The authors achieved this versatility by giving both two-point and three-point field goals the appropriate weight for their respective shooting percentages. In basketball, there are several situations similar to those mentioned above in which a team would greatly benefit from looking at the WEFG, as this value gives more detailed game information. In basketball, the team with the most points wins. However, if you break the game down further, it is comprised of a series of smaller games. Using statistical analyses, it can be determined how winning each one these games helps increase the teams' chances of ultimately winning. Understanding the strengths and weaknesses of a team will allows a team to have the most effective strategy and to better prepare for games.

Hockey Analytics

Analytical models first used in hockey primarily focused on goal-based metrics, such as the plus-minus analysis (Found, 2016). While the plus-minus analysis was first introduced in hockey, it has become a standard for most professional sports and has be altered as needed as better analysis required the process to evolve (Macdonald, 2012). The traditional plus/minus statistic is a simple point system. Players are awarded a positive-point if they are on the ice at the time their team scores a goal. Players receive a negative-

point if they are on the ice at the time a goal was scored against their team. The advancement of various analytical models in hockey has led sport professionals to rely more and more on shot-based metrics. The Corsi value has become quite prominent in recent years. Corsi was named after Buffalo Sabres' goalie coach, Jim Corsi. The original idea behind the Corsi value was to track the workload for goalies, as the value looks at not only shots on goal, but attempted shots and blocked shots as well (Vollman, 2016). For each of these situations, the goalie would be anticipating the shot.

The Corsi value has essentially become a glorified plus-minus statistic. Instead of having the focus on points, the focus is on attempted shots. The NHL has started to track individual Corsi values, but these values have been renamed as shot attempts (SAT). The NHL tracks SAT by taking the sum of the shots on the opposing goal (while that individual is on the ice), subtracted by the shots taken on their own net (while that individual is on the ice). As with the plus-minus statistic, anything above zero is a positive outcome and anything below zero is negative. Corsi value at the team-level is calculated by: $\frac{Shot\ Attempts\ For}{Shot\ Attempts\ For + Shot\ Attempts\ Against}$. For this value, fifty percent is where an average team falls. The higher the number, the better the team is performing, at least when it comes to attempted shots. A common criticism of the traditional Corsi measurement is that, when looking at the individual-level, players have the tendency to take shots from anywhere on the ice. The rationale behind this is that a player could intentionally increase their individual Corsi value, which would increase their value to their respective team. This can create quite a conflict for team management. Found (2016) examined this quandary and concluded that goal-based metrics (plus-minus) were better indicators of individual and team success than using Corsi because many of the shots taken were not effective enough to potentially result in goals.

When evaluating individuals, it is important to take the veracity of their Corsi value into account; that is why Corsi needs to have weighting. While remembering that Corsi was developed to focus on the goalie, placing both the save-average and goals-against into the weighted Corsi value increases the accuracy of the value. The formula is relatively simple $\frac{Shots\ For}{Shots\ For\ +\ (\frac{Goals\ Against}{1-Save\ Percentage})}$ the save

percentage is calculated by $\frac{Shots\ Against - Goals\ Against}{Shots\ Against}$. The reason why goals-against and save-percentage are used in this equation is that they are not only key goaltending statistics, but they also include a value from the traditional Corsi value: shots-against.

The use of Modified Corsi is supported by the statistical analysis of data from ten recent NHL seasons, from 2008 to 2017. This includes every goaltender that has stepped onto the ice during that time. The mathematical backing is comprised of over 3,000 data points. In projecting the top sixteen teams in the NHL, without taking divisions into account, the traditional Corsi value was able to project, on average, eight of the sixteen teams. Conversely, the Modified Corsi was able to project, on average, twelve of the sixteen top teams. That results in a twenty-five percent increase when projecting the top 16 teams, as compared to the traditional Corsi value. When looking at the league, the Modified Corsi value was consistently able to project fourteen percent more accurately than the traditional Corsi value. The Modified Corsi value had an average of eighty percent accuracy in projecting the top teams over the past ten seasons. Figure 1 portrays the percentage of teams accurately projected to make the playoffs each year from 2008 to 2017, using both traditional and Modified Corsi.

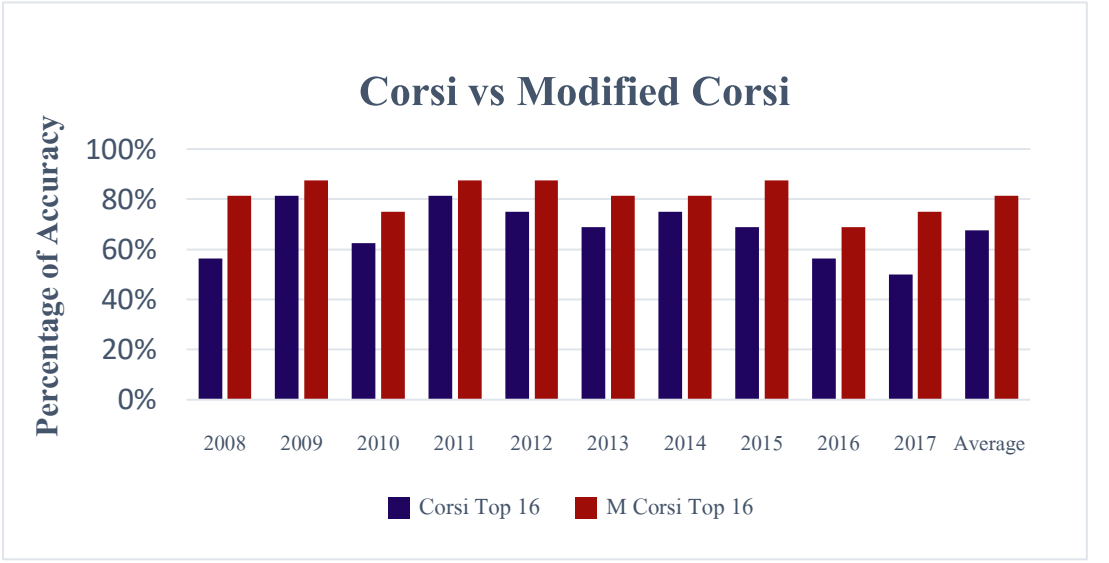

Figure 1. Corsi top 16 NHL teams compared to Modified Corsi top 16 NHL teams.

Using this Modified Corsi value can be extremely valuable to an organization. Calculating the numbers on these modified Corsi values at the beginning of the season and projecting these numbers to the end of the season, organizations would be able to adjust a variety of areas, based on their initial results. Coaching staff and team management regularly need to find answers to the following questions: how is goaltending performing, is there enough blocking of shots, and is the offence or defense taking enough shots or preventing enough shots? With Modified Corsi, there are key factors that can significantly change its overall value. These factors come from every area of team performance, which is why this value has is such an accurate projection for the top teams in the league.

Conclusion

There is evidence to suggest that traditional analytic methods do not always obtain the most accurate results (Found, 2016; Macdonald, 2012). For this reason, the advancement of analytics continues to evolve as sport professionals strive to understand team performance and the value of individual players value to their teams. The models presented may be useful to analysts, coaches, teams and fans, as they continue to analyze the performance of individuals and teams. However, only focusing on the statistical analysis may cause a loss of direction for building an effective team. When these analytical methods are used, there may not be a symbiotic relationship between analysis, coaching staff, and management. Furthermore, while analytics is one tool that is gaining prominence in the sport industry, it should not be the only standard measured, as it is easy to get lost in copious data. There needs to be a firm understanding of the game and having a clear direction is the best way for an organization to achieve success. The only way to accomplish this is to have an open line of communication from the top down. So long as all parties within the organization are on the same page, the results of these analytics can be more effectively applied. Utilizing the weighted formulas discussed in this paper provides additional tools that an organization can use to help determine the direction for their team.

References

Bedford, A. & Baglin, J. (2009). Evaluating the performance of an ice hockey team using interactive phases of play. IMA Journal of Management Mathematics, 20, 159-166

Chan, T. C. Y., Chow, J. A. & Novati, D. C. (2012). Quantifying the contribution of NHL player types to team performance. Interfaces, 42, 131-145.

Cochran, J. J., & Blackstock, R. (2009). Pythagoras and the National Hockey League. Journal of Quantitative Analysis of Sports, 5(2). DOI https://doi.org/10.2202/1559-0410.1181

Found, R. (2016). Goal-based metrics better than shot-based metrics at predicting hockey success. The Sport Journal, 20.

Hamilton, H. H., (2011). An extension of the Pythagorean expectation for association football. Journal of Quantitative Analysis in Sports, 7(2). DOI: https://doi.org/10.2202/1559-0410.1335.

Horowitz, I. (1994). Pythagoras, Tommy Lasorda, and me: On evaluating baseball managers. Social Science Quarterly, 75(1), 187-194.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. (2007). A starting point for analyzing basketball statistics. Journal of Quantitative Analysis in Sports, 3(3). DOI: https://doi.org/10.2202/1559-0410.1070.

Macdonald, B. (2012). Adjusted plus-minus for NHL players using ridge regression with goals, shots, fenwick, and corsi. Journal of Quantitative Analysis in Sports, 8(3). https:// DOI: 10.1515/1559-0410.1447

Moskowitz, T., & Wertheim, J. L. (2012). Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won. Crown Publishing Group.

Vollman, R. (2016). Stat Shot. Toronto, ON: ECW Press.