Forecasting NHL Hat Tricks

Sam Forstner

Abstract

In this paper, I explain the frequency of hat tricks in the National Hockey League (NHL). Assuming goal scoring follows a Poisson process, I propose models intended to generate an accurate number of hat tricks across multiple seasons. Additionally, I examine whether the players who score hat tricks are typically the ones we would expect to do so, given rates of goal scoring. I use ice time and goal scoring data from HockeyReference.com from the 2000-2001 through 2016-2017 seasons and construct three models for my analysis—one that pools over players and seasons, one that allows for season-to-season changes, and one that allows for different scoring rates by player. Finally, I estimate the expected number of hat tricks scored by the highest scoring players of the years 2000-2017 and compare estimates to the observed numbers scored.

1. Introduction

This paper develops an application for Poisson random variables and applies it to hockey. Additionally, the paper provides insight into not just the sport of hockey but the occurrence of extreme events in sports more generally. Increasing amounts of time and money are being invested in attempting to quantify and analyze various aspects of sports every year—for example, in 2016 the NHL's Arizona Coyotes hired 26-year-old John Chayka, the co-founder of a hockey analytics firm, to be their general manager (Lago, 2016). This paper attempts to broaden that sports analytics knowledge base. Finally, the conclusions of this paper could prove interesting to sports viewers, as extreme events like hat tricks that appear to be the result of extraordinary circumstances could be shown to be somewhat predictable.

The paper proceeds as follows. Section 2 gives relevant background information and grounds the paper in the existing literature. Section 3 describes the data and outlines the methods used. Section 4 presents results. Finally, Section 5 provides conclusions and discussion of the findings.

2. Background and Significance

In this project, I attempt to model the occurrence of hat tricks in the NHL, assuming goals occur as a Poisson process. I will first explain some basic terminology that is central to the project.

2.1: What is a "Hat Trick"?

A "hat trick" in hockey occurs when a single player scores three or more goals in one game. These events are rare. There were 59 hat tricks in the NHL in the 2016-17 season in 1,230 total games, meaning hat tricks occurred only about once in every 20 games.

2.2: The Poisson Distribution

The Poisson Distribution expresses the probability of a given number of events occurring in a certain interval of time, given that the events occur with a constant rate and are independent of the time since the last event—i.e., the events are memoryless.

The Poisson Distribution P (X = k) with parameter λ is the continuous limit of the binomial distribution b(n,p,k) with pm = λ as $n \to \infty$ (Krishnamoorthy, 2016).

$$P(X = k) = \lim_{n\to\infty} b(n, \frac{\lambda}{n}, k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

In the context of an NHL hockey game, the interval of time is 60 minutes, the rate λ is the average number of goals scored per 60 minutes of play, and k represents a number of goals in that interval. For example, if a player scores on average 0.5 goals per 60 minutes, the probability of him scoring three or more goals in that amount of time is approximately 0.01.

2.3: Goal Scoring in Hockey and Soccer as Poisson

Ryder (2004) explores how goal scoring in hockey can be explained by Poisson models, and he shows how goals in hockey fulfill all of the requirements of a Poisson process. His work provides evidence that goal scoring is Poisson, and provides the example of a distribution of NHL shutouts induced by this Poisson model for goal scoring.

Mullet (1977) models NHL goal scoring as a Poisson process, generating goal-scoring factors for each team for both home and away games. He uses this model to attempt to produce accurate NHL standings for the 1973-74 season, finding that, for most teams, their forecasted and actual positions in the standings were not significantly different.

Buttrey et al. (2011) model NHL goals (goals for and goals against for each team) as Poisson processes. The rate in their model depends on the two teams playing, which team has home-ice advantage, and the manpower situation (if one team has more players on the ice than the other due to penalties). The authors compute offensive and defensive factors for each team and construct a goals-by-manpower dataset, reflecting the average rate at which goals are scored in each manpower situation (five players vs. three, for example). Buttrey et al. then construct a Poisson generalized linear model (GLM) for any manpower situation between any two teams, and employ this model to compute team win probabilities and attempt to replicate the NHL standings for the 2008-2009 season. The authors test the fit of their model by employing cross-validation and comparing the predictions to the actual standings.

Goal scoring in soccer has also been studied as a Poisson process. Dyte and Clarke (2000) use a Poisson model for goals to simulate World Cup games, predicting a near-exact number of total goals in the tournament. Chu (2003) uses World Cup data to construct a Poisson model for goal arrivals, finding that the actual number of goals scored in given time intervals and the number predicted by the model are not significantly different from one another. Goddard (2005) uses a bivariate Poisson regression for English soccer goal scoring to forecast match results, finding that the model performs similarly to a results-based probit model. Lee (1997) uses a Poisson model for goal scoring to question whether Manchester United was truly the best Premier League team in the 1995-96 season, concluding that they likely were. Ridder et al. (1994) measure the effect of a red card (being ejected from a game) on Poisson "scoring intensity," finding that this intensity increases 88% for the team that still has 11 players as opposed to ten.

2.4: Extreme Events in Sports

Sichel et al. (2011) attempt to explain the occurrence of perfect games and no-hitters in baseball using simple mathematical models, and examine whether the pitchers who have thrown perfect games are the ones that one would "expect" to do so. The authors compute the probability of a perfect game occurring using the probability of an individual runner reaching base, the idea being that 27 consecutive runners not reaching base results in a perfect game. They develop three models: a simple model in which all players and seasons are treated equally, a season-by-season model, and finally a pitcher-by-pitcher model. This most sophisticated model produced a number of perfect games and no hitters within one standard deviation of the amount that actually occurred. Huber (2007) analyzes extreme events in baseball—no-hitters, hitting for the cycle, and triple plays—and models these events as Poisson processes.

2.5: My Contribution

This project aims to contribute to the existing literature in two primary ways:

(1) Expand the application of Poisson processes to the sport of hockey

(2) Provide additional insight into modeling the occurrence of extreme events in sports

My contribution builds on previous work with hockey and soccer scoring as a Poisson process while employing an analytical framework similar to that of Sichel et al.'s study of extreme events in baseball.

3. Data and Methods

3.1: Data

I use a dataset of NHL player statistics spanning the 2000-2001 through 2016-2017 seasons from HockeyReference.com. I chose this range as there were a consistent 30 teams in the league during this time and thus a constant number of games per (complete) season. I limit my sample to forwards (I exclude defensemen) under the assumption that goal scoring is fundamentally a function of position; defensemen score significantly less than forwards, so treating time played by defensemen equally in the forecasting of hat tricks would not lead to an accurate prediction. In other words, there is a goal scoring rate for forwards, defensemen, and goalies (though this would be miniscule), and I am only interested in modeling the forward rate.

The independent variables for my analysis are player ice time (measured in seconds by game) and player goals scored, which are used to produce the paramater estimate λ. The dependent variable of interest is the observed number of hat tricks scored in a season.

3.2 Design

3.2.1: Hypotheses

I test two hypotheses:

H1: A Poisson model for goal scoring generates an accurate number of hat tricks in an NHL season

H2: The players with the highest goal-scoring rates score the most hat tricks

I employ three models to test my first hypothesis, ranging from a naïve model that treats all players and seasons the same to a more sophisticated model where each player has his own scoring rate parameter λ.

3.2.2: Pooled Model

In the pooled model, I compute one league-wide (pooled) goal-scoring rate by dividing the total goals scored over a period by the total number of minutes played, then multiplying by 60 to scale it in terms of "player-games".

$$\lambda_{pooled} = 60 * \frac{G_{pooled}}{TOI_{pooled}}$$

Note: Even in this pooled model, λ is still scaled by each individual player's ice time (as a fraction of 60 minutes) in each game to generate an expected number of hat tricks for that player in that game. I outline the process of generating these expected values in detail in Section 3.2.5.

3.2.3: Season-by-Season Model

The methodology for the season-by-season model is similar to the pooled specification, except that all statistics feeding into the model are calculated for each individual season, producing a rate parameter λ_t where $t =$ season for each season as shown below.

$$\lambda_t = 60 * \frac{G_t}{TOI_t}$$

3.2.4: Player-by-Player by Season Model

The player-by-player model follows the same progression and compute a λ for each player in each season, producing hundreds of rate parameters per season (λ_{ti}), where $i =$ player and $t =$ season .

$$\lambda_{ti} = 60 * \frac{G_{ti}}{TOI_{ti}}$$

3.2.5: Forecasting Hat Tricks

For each model, I calculate the expected number of hat tricks in each "player-game" by using the corresponding λ. The probability of scoring at least three goals in a game is equal to the probability of not scoring two, one, or zero goals, so:

$$P(X > 3) = 1 - (e^{-\lambda}\frac{\lambda^2}{2!} + e^{-\lambda}\frac{\lambda^1}{1!} + e^{-\lambda}\frac{\lambda^0}{0!})$$

where X = number of goals scored.

The 60-minute rate calculated for each model is scaled for each player-game before calculating the probability above. For example, if a player plays 15 minutes in a game, the expected number of hat tricks scored by that player in that game under the pooled model is calculated using $\lambda = \frac{\lambda_{pooled}}{4}$.

Finally, the expected number of hat tricks in each player-game over a given period is summed and compared to the observed number of hat tricks scored in that period.

I test the models in a few ways. First, if forecasting over multiple seasons, I calculate the root mean squared error of the estimates for each season to assess their accuracy. Second, I employ cross validation by splitting the season in half and attempting to "predict" the number of hat tricks in the second half of a season using a model constructed from data from the first half. Finally, I compare this approach to "naive" guessing to further evaluate its accuracy.

Finally, to test my second hypothesis, I compile a list of 56 players with the highest expected number of hat tricks they would have scored over the period 2000-2017, comparing these predictions to the observed results. I also evaluate this hypothesis using root mean squared error.

4. Results

4.1: Summary Statistics

In order to forecast hat tricks by season, I first generate a rate of goal scoring for each season using the total number of minutes played by each player and the total number of goals scored by season. This λ represents the average "player-goals" per minute over the entire season. I then scale this λ by the average number of minutes played per game over this period, to create a goals-per-game rate for the "typical" NHL forward in each season. Summary statistics of these scaled estimates of λ are shown in Table 1.

	Min	Mean	Max
1	0.1830	0.1949	0.2207

Table 1: Summary of λ by Season

The average number of minutes played over this period was approximately 15 minutes per game per player. These estimates represent the average number of goals scored per game by a fictional "typical" player who played on average that amount in that season. So, a player who played 15 minutes per game during this period scored on average approximately 0.19 goals per game.

Next, I produce estimates of λ for each individual player for each season, using that player's total time on ice and number of goals scored. I again scale these estimates of λ by the average number of minutes played by forwards over this time to frame the rates in terms of goals-per-game for a "typical" NHL game. Thus, each player's λ reflects approximately how many goals that player would score per game playing the average number of minutes. The Top 20 players by this single-season estimated λ are shown in Table 2 (Appendix). Joffrey Lupul's 2012-2013 season with Toronto tops the list, as he scored 11 goals in just 16 games while averaging 16:07 of playing time per game.

Figure 1 shows the distribution of player λ estimates, with the estimates on the horizontal axis and the number of player-seasons exhibiting that estimate of λ on the vertical axis. As shown, the distribution is extremely right-skewed. Most players cluster around the mean of 0.19 or at 0, as many players did not score at all, while top performers like Lupul produce estimates of λ of 0.63 goals per "typical" game.

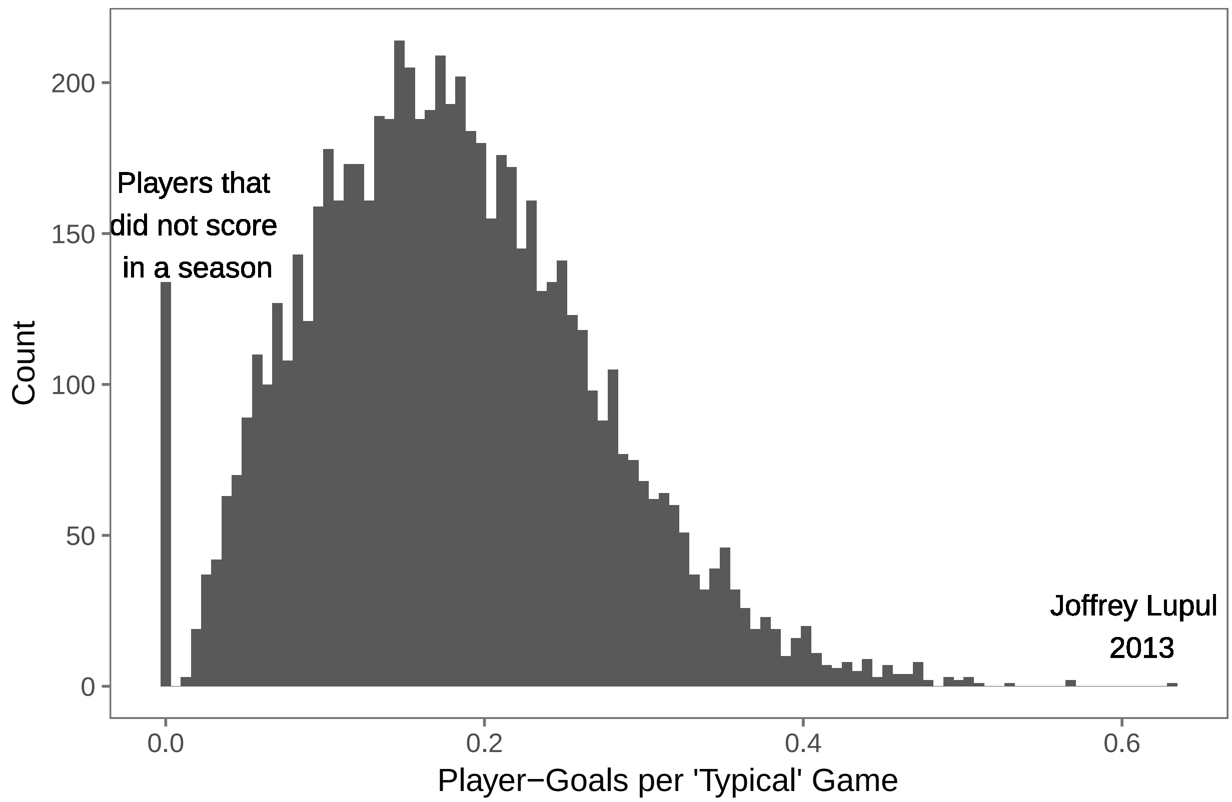

Figure 1: Distribution of Lambda by Player (by Season)*

Note: Given that a player played at least 200 minutes in that season. Some players exhibit extremely high estimates of λ over a small sample size. Samuel Henley, who played just one game in 2016-17 for the Colorado Avalanche, is the most extreme case of this, as he scored one goal in just over five minutes of playing time for an estimated λ of 2.8 goals per "typical" game. As this was the only NHL game Henley has ever played, it would clearly not be appropriate to include his goal-scoring rate of almost one hat trick *per game* in this distribution.

4.2: Forecasting Hat Tricks by Season

First, I test my primary hypothesis that a Poisson model for goal scoring can accurately predict the number of hat tricks in an NHL season.

4.2.1: Pooled Model

I first forecast the number of hat tricks pooling across all players and all seasons. I calculate a single λ using the total number of goals scored and total time on ice in this 17-year period, relying on the likely naive assumption that all players can be treated eaually. This approach leads to an estimate of 608.23, when in reality, 1020 hat tricks occured. This model underestimates the total number of hat tricks, and clearly a better approach is needed.

4.2.2: Season-by-Season Model

To attempt to improve upon the pooled model, I propose a model that accounts for variation between seasons. This model uses a λ for each season to produce an expected number of hat tricks for that season. The results of this model are shown in Table 3 (Appendix).

This approach leads to a total estimate over 17 seasons of 612.73 hat tricks, not far from the previous pooled estimate of 608.23. However, I do not find this lack of improvement from the pooled model to the model that allows for variation by season surprising, as the temporal variation in goal scoring between even the highest and lowest scoring seasons is small (as shown in Table 1).

To evaluate the accuacy of this model, I compute its Root Mean Squared Error (RMSE). The Season-by-Season model produces an RMSE of 27.24 hat tricks.

4.2.3: Player-by-Player Model

Next, I account for variation between players, generating a λ for each player in each season using that player's total goals scored and time on ice in each season. The complete results of this model are shown in Table 4 (Appendix).

This model marks an improvement over prior specifications. It produces a total estimate over 17 seasons of 1081.85, close to the actual number of 1020.

Further, the player-by-player model improves the RMSE to 8.83, a 208.5% improvement. The performance of the model on a player-by-player level over the 17-year period is shown below in Figure 2. Each point represents a player, with forecasted hat tricks on the horizontal axis plotted against observed hat tricks scored on the vertical axis. The shading of a point corresponds to the player's ice time, with darker shading indicating more time on ice.

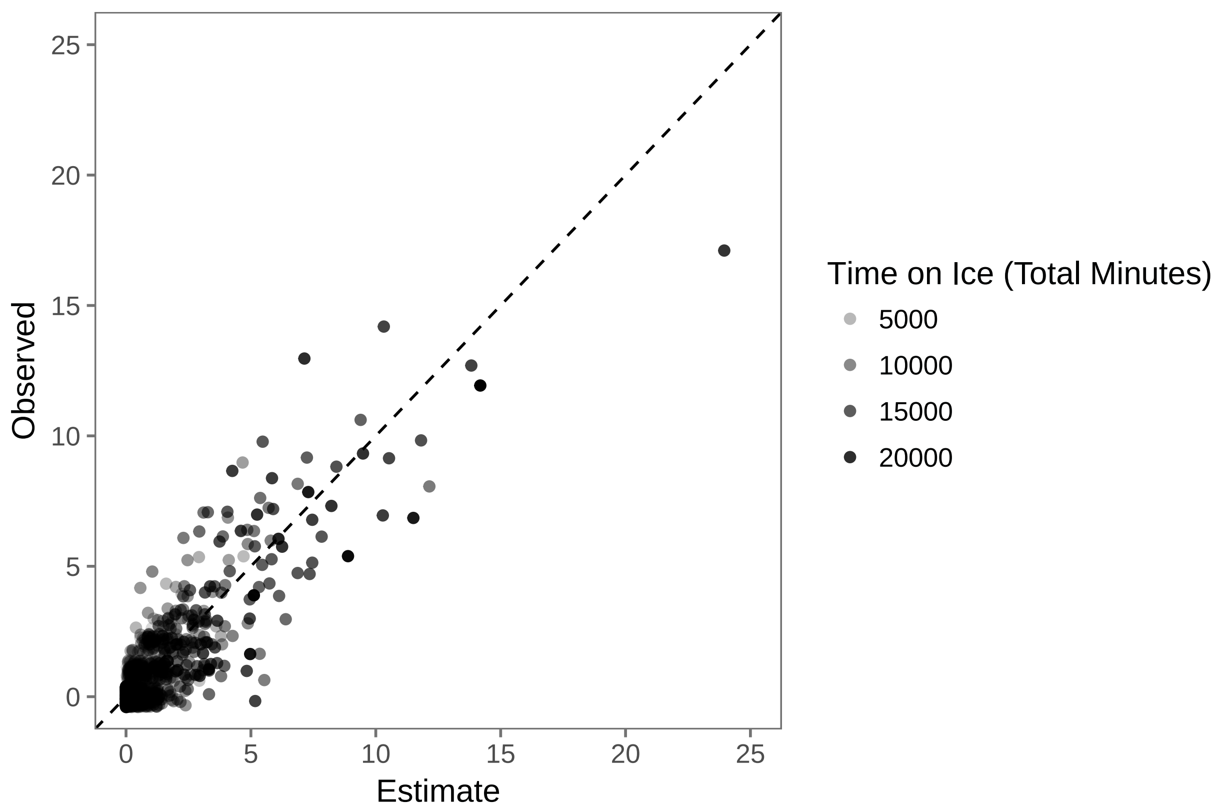

Figure 2: Forecasted and Observed Hat Tricks by Player

4.2.4: Model Comparison

The results of both the Season-by-Season and player-by-player specifications are shown in Table 5 (Appendix).

The performance of the season-by-season and player-by-player models can be seen in Figure 3. Both plots show forecasted hat tricks on the horizontal axis and observed hat tricks on the vertical axis, with the dotted lines representing a one-to-one linear relationship. Note the consistent underestimation of the season-by-season model as compared to the approximately linear relationship shown in the player-by-player specification.

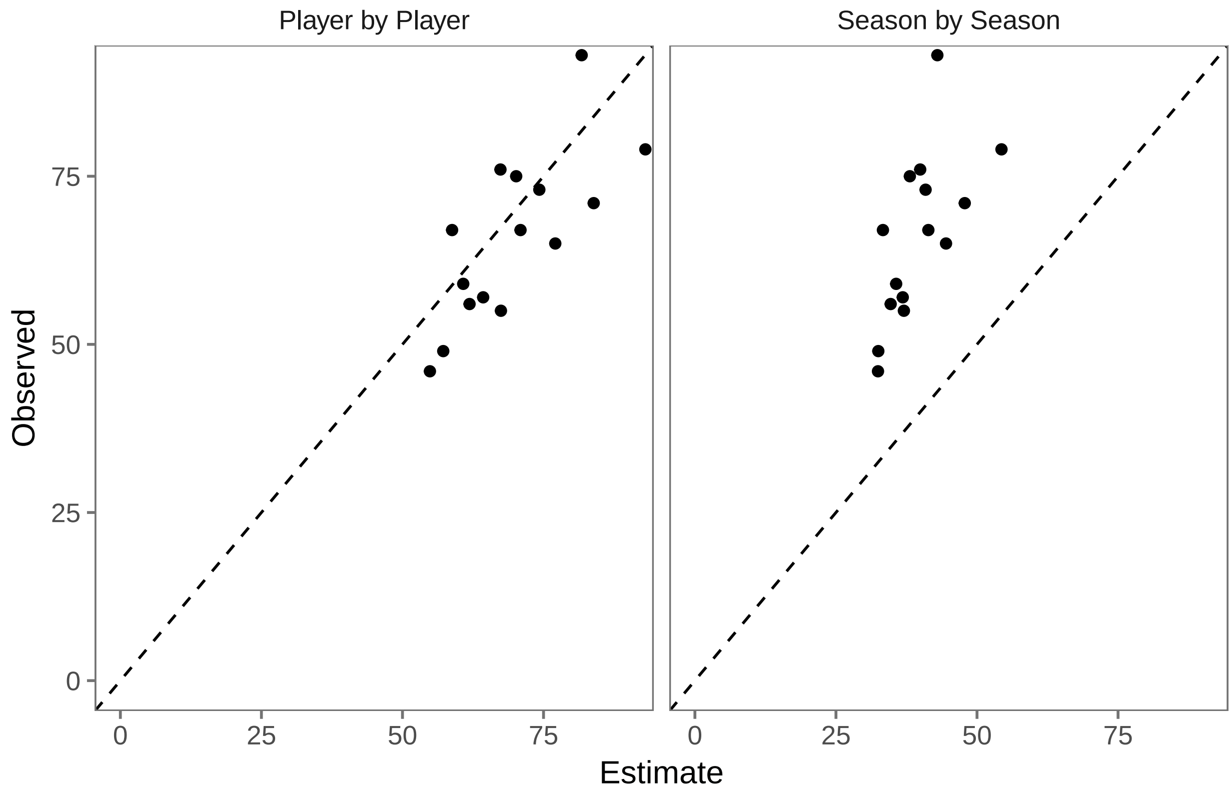

Figure 3: Forecasted and Observed Hat Tricks by Season (2001−2017)

Note: The 2012-2013 season was shortened by a lockout, and teams played only 48 games instead of the usual 82. As a result, that season has been dropped from this analysis.

4.2.5: Cross-Validation of Player-by-Player Model

Though the player-by-player model forecasts hat tricks accurately, one potential issue is that of overfitting. That is, does this model just fit the specific player-by-player dataset without being applicable to forecasting hat tricks more generally? One way to check for this problem is through cross-validation. I split my sample in half, using January 31 as the cutoff for each season. I selected this date for the cutoff because the NHL All Star Game usually occurs in late January; and the All Star Game is played at roughly the halfway point of the season. Then, I use the same methodology as above to produce estimates of λ using only the data from the first half of each season, and attempt use those estimates of λ to forecast hat tricks in the second half of each season. The results of this test are shown visually on the season and player level in Figures 4 and 5, respectively, and in Table 6 (Appendix).

This method of forecasting is highly sensitive to anomalies in which a player scores at a high rate in a small amount of ice time in the first half of the season and then plays much more in the second half. In situations like this, disproportionate weight is given to a small sample from the first half when it is applied to the much larger number of minutes in the second half. To attempt to account for this, I include only observations in which the player played at least as many minutes in the first half as in the second. Additionally, I once again drop the lockout-shortened 2012-2013 season from this analysis.

I evaluate this method using RMSE. This application of cross-validation yields an RMSE of 8.24 hat tricks.

Figure 4 shows the results of the twofold cross-validation, with forecasted hat tricks on the horizontal axis and observed hat tricks on the vertical axis and each point representing one season.

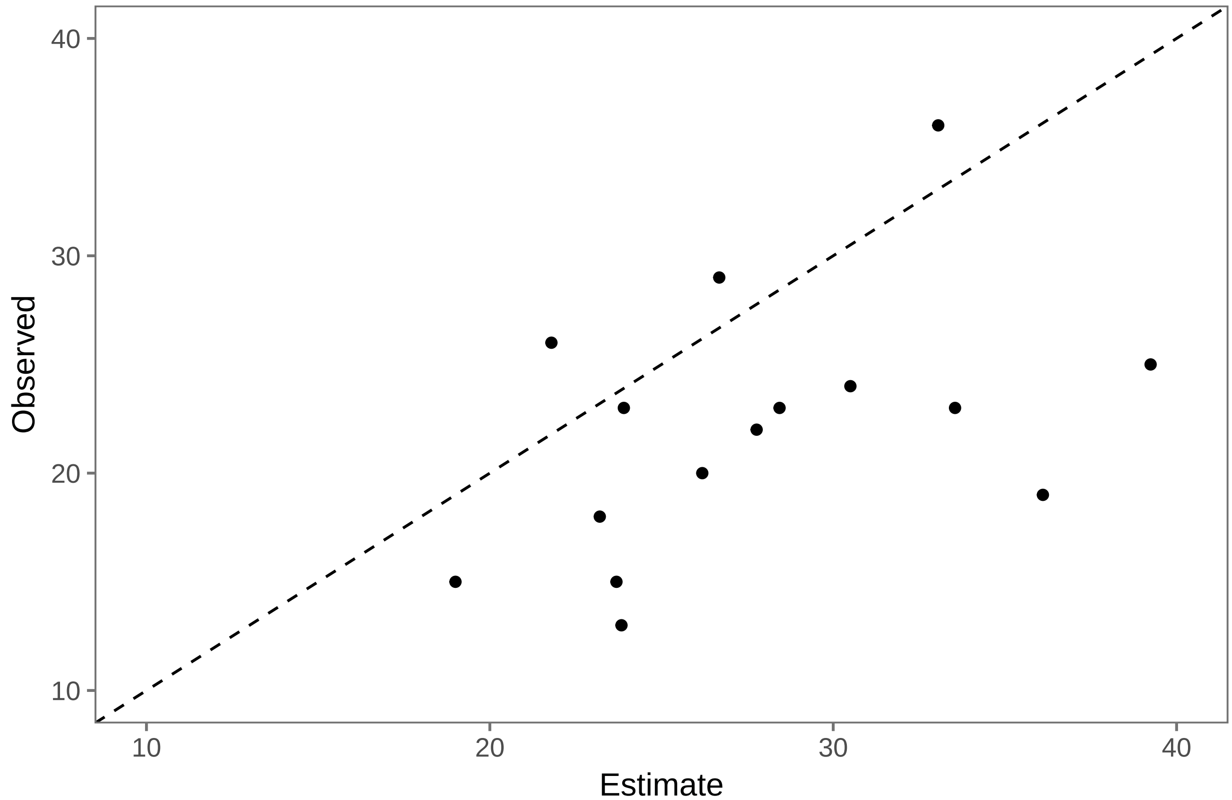

Figure 4: Forecasted and Observed Hat Tricks in Second Half of Season

Figure 5 again shows the relationship between forecasted and observed hat tricks in the second half of the season, this time at the player level. The shading of the points represents player ice time, with darker shading indicating more playing time.

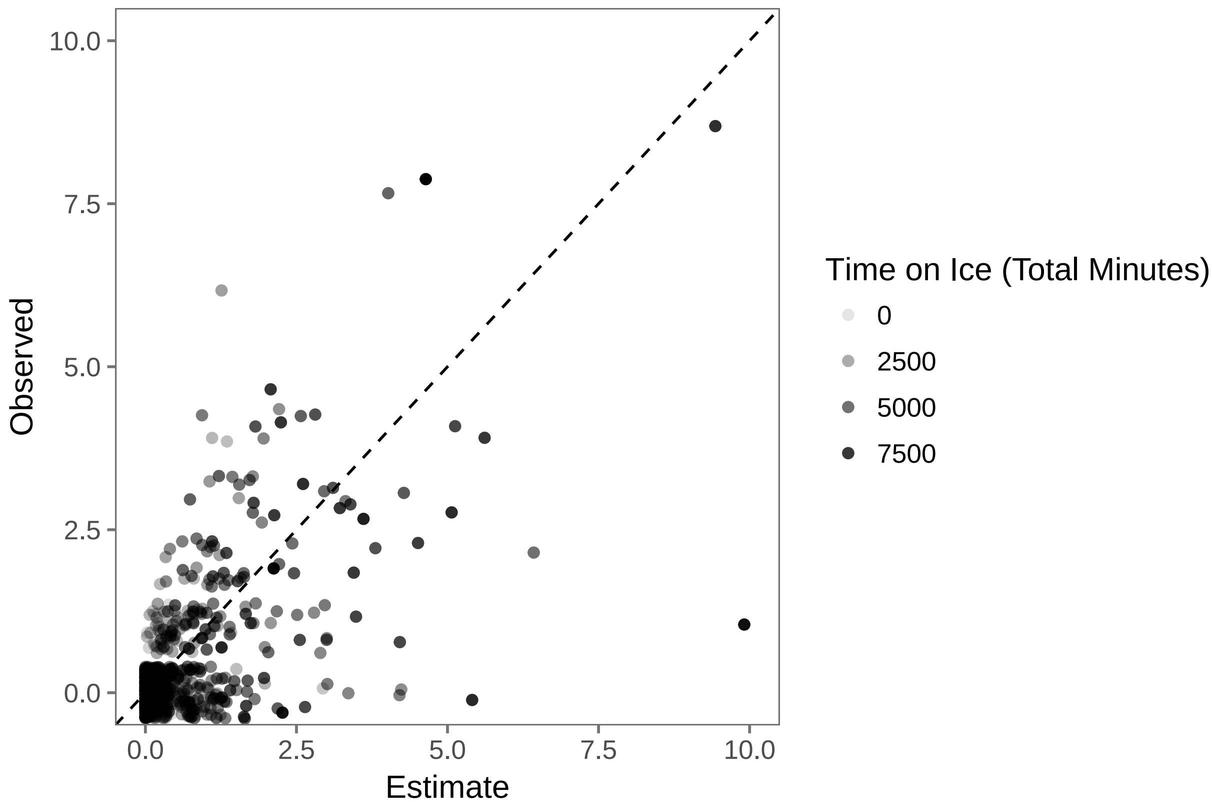

Figure 5: Second Half Forecasted and Observed Hat Tricks by Player

4.2.6: Does Model Beat Naive Guessing?

It is difficult to truly judge the performance of this model without comparing it to an alternate method of forecasting hat tricks. Here I make that comparison with a naive model that simply assumes that the 1020 total hat tricks scored between the 2000-01 and 2016-17 seasons are uniformly distributed. A comparison between the cross-validation and "naive" forecasts is shown below in Figure 6. Both plots show forecasted hat tricks on the horizontal axis and observed hat tricks on the vertical axis, with the dotted line representing a linear relationship for context. Each point represents one season.

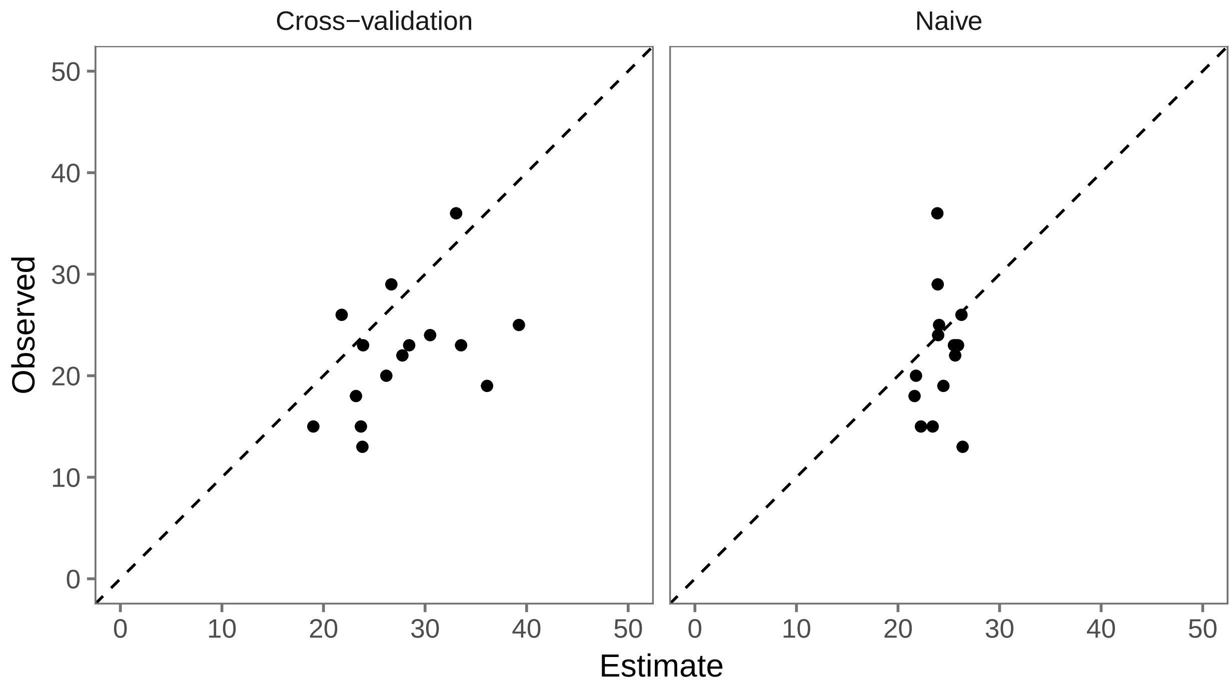

Figure 6: Second−Half Forecasted and Observed Hat Tricks
(Player−by−Player and Naive Guessing)

As shown above, cross-validation produces an approximately linear relationship between forecasted and observed hat tricks, while the naive assumption that hat tricks are uniformly distributed produces estimates that are clustered together. However, from an RMSE perspective this naive guessing is actually *more accurate*, producing a RMSE of 6.09, 35.2% better than twofold cross-validation.

I believe that this is not surprising given that the method of cross-validation is highly sensitive to anomalies in the first-half λ, and it is possible that the accuracy of some of the naive guesses is the result of chance. Although I attempted to account for these anomalies, I expect that variation between individuals' first and second-half performance attenuate the effectiveness of this split-sample approach.

4.3: Forecasting Hat Tricks for Top 56 Players

To test my second hypothesis, that the players who score hat tricks are the ones we would expect to do so, I focus on the top 56 players by expected number of hat tricks (using the player-by-player specification). I compare their expected number of hat tricks over the period of 2001-2017 to the observed number of hat tricks each player scored, as shown in Table 7 (Appendix).

I use RMSE to evaluate the accuracy of forecasting for the top 56 players, as well. This prediction yields an RMSE of 2.61 hat tricks, suggesting that the players we expect to score hat tricks are in fact the ones doing so.

The relationship between forecasted and observed hat tricks for these players is shown below in Figure 7, with forecasted hat tricks on the horizontal axis and observed hat tricks on the vertical axis.

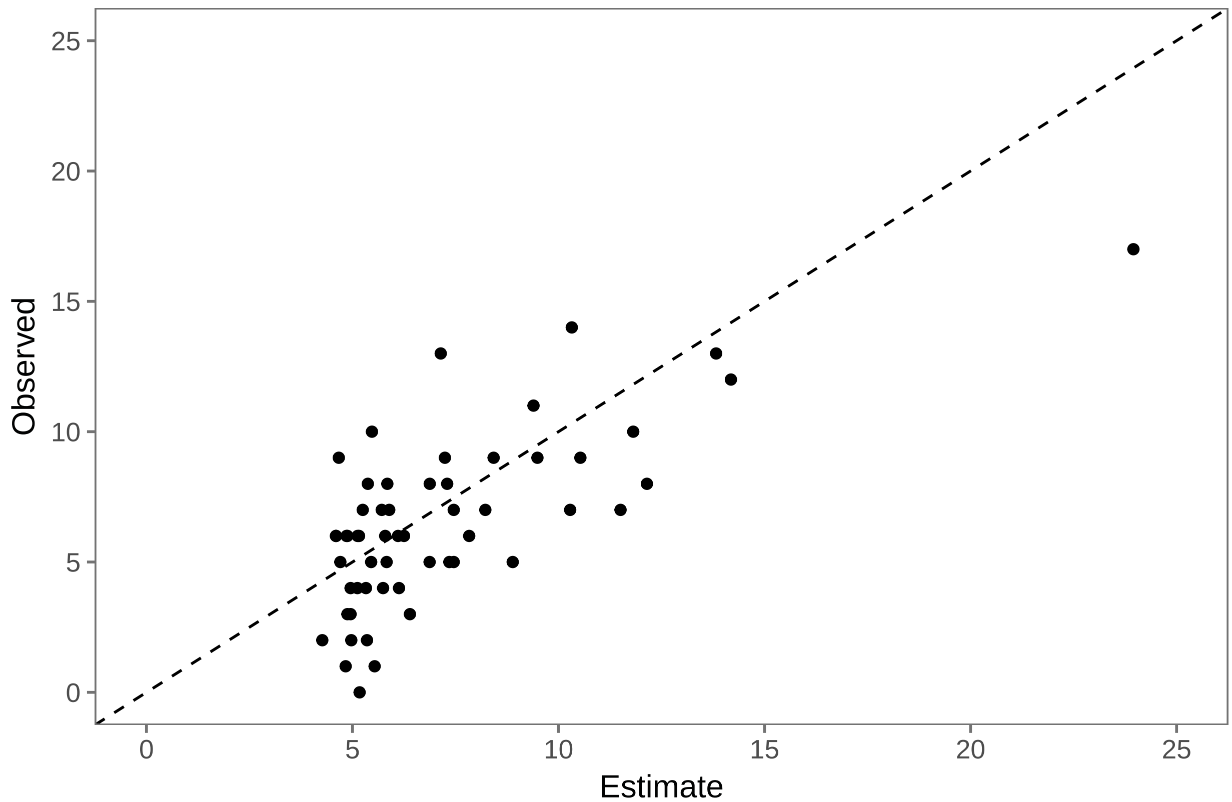

Figure 7: Top 56 Players in Expected Hat Tricks (2001−2017)

Finally, I check whether the players who score the most hat tricks are the ones we expect to do so by cross-referencing the list of top players by expected hat tricks with a list of the top players by observed hat tricks scored. This is how I arrived at 56 for the number of players; 56 players scored 5 or more hat tricks over this period, so using only the top 50 would exclude some players who had actually scored just as many hat tricks as those on the list. Of the 56 players who scored 5 or more hat tricks, 42 of them are in the top 56 in expected hat tricks as well. The results of this comparison are shown in Table 8 (Appendix).

5. Conclusion and Discussion

I have constructed a dataset of ice time and goal scoring for all NHL forwards from 2000 to 2017, and used this to predict the occurence of hat tricks over this period. By assuming goal scoring is Poisson and using a model that allows for variation in goal-scoring rates between players and between seasons, I am able to forecast hat tricks accurately. Additionally, I have found that the players who score the most on average do score the most hat tricks, and that most (75%) of the players who scored the most hat tricks were the ones I expected to under the forecasting model.

The forecasting ability of the player-by-player Poisson model is strong evidence of hockey goal scoring being a Poisson process. However, there are some limitations to this study.

First, there is the issue of endogeneity. Player ice time is likely a function of performance—i.e., a player's likelihood of scoring a hat trick. In other words, if a player scores more, he is likely to receive more ice time. So, because I use player goal scoring and player ice time to generate a player's goal-scoring rate λ and then calculate the expected number of hat tricks in each game by scaling that λ to the player's ice time in that game, my predictions are likely biased upwards.

Second, though this model is useful for analyzing and forecasting hat tricks *within* individual seasons, it is not meant to predict hat tricks in future seasons. The predictive power of the model is limited by the amount of variation there is between seasons. The players are not constant—they enter and exit the league, change teams, and gain and lose ability with age and injury. Additionally, league rules change frequently and impact scoring, further limiting the ability of these models to predict hat tricks in one season given scoring in a previous one.

The issue of ice time's endogeneity presents interesting questions for further study. I would be interested in how a player's ice time changes in an individual game when he has already scored two goals. Additionally, I am curious if players' behavior changes when they have scored twice and are "on the verge" of a hat trick. Do they shoot more, pass less, or make more mistakes as they push for the milestone? I believe this would be an interesting question to investigate using individual game log data.

I am also interested in the 14 players who scored five or more hat tricks but were not predicted to by my model. Does this group of players have any unifying characteristics? Are they more likely to score "in bunches" (more than once in a game) when they do score, or exhibit "streaky" behavior? I believe these questions are also worth investigating.

References

Buttrey, S. E., Washburn, A. R., & Price, W. L. (2011). Estimating NHL scoring rates. *Journal of Quantitative Analysis in Sports,*7(3).

Chu, S. (2003). Using soccer goals to motivate the Poisson process. *INFORMS Transactions on Education,*3(2), 64-70.

Dyte, D., & Clarke, S. R. (2000). A ratings based Poisson model for World Cup soccer simulation. *Journal of the Operational Research society,*51(8), 993-998.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of forecasting,*21(2), 331-340.

Huber, M., & Glen, A. (2007). Modeling Rare Baseball Events–Are They Memoryless?. *Journal of Statistics Education,*15(1), n1.

Krishnamoorthy, K. (2016). *Handbook of statistical distributions with applications.* CRC Press.

Lago, K. (2016, November 1). *How the youngest general manager in professional sports history got his job.* Retrieved from https://www.cnbc.com/2016/11/01/how-the-youngest-general-manager-in-professional-sports-history-got-his-job.html

Lee, A. J. (1997). Modeling scores in the Premier League: is Manchester United really the best?. *Chance,*10(1), 15-19.

Mullet, G. M. (1977). Simeon Poisson and the National Hockey League. *The American Statistician,*31(1), 8-12.

Ridder, G., Cramer, J. S., & Hopstaken, P. (1994). Down to ten: Estimating the effect of a red card in soccer. *Journal of the American Statistical Association,*89(427), 1124-1127.

Ryder, A. Toolbox: a review of the application of the Poisson probability distribution in hockey. 2004. www.hockeyanalytics.com

Sichel, R., Carl, U., & Bukiet, B. (n.d.). Modeling Perfect Games and No-Hitters in Baseball. 2011. http://sabr.org/research/modeling-perfect-games-and-no-hitters-baseball

Appendix

	Name (career span, position)	Season	Lambda
1	Joffrey Lupul (2004-2016, LW)	2013	0.6315
2	Marian Gaborik (2001-2018, RW)	2009	0.5664
3	Marek Svatos (2004-2011, RW)	2006	0.5652
4	Sidney Crosby (2006-2018, C)	2011	0.5273
5	Alex Ovechkin (2006-2018, LW)	2008	0.5081
6	Jonathan Cheechoo (2003-2010, RW)	2006	0.5067
7	Alexander Mogilny (1990-2006, RW)	2001	0.5029
8	Daniel Briere (1998-2015, C)	2001	0.5009
9	Adam Deadmarsh (1995-2003, RW)	2003	0.4987
10	Mario Lemieux (1985-2006, C)	2001	0.4953
11	Steven Stamkos (2009-2018, C)	2014	0.4941
12	Steven Stamkos (2009-2018, C)	2012	0.4921
13	Teemu Selanne (1993-2014, RW)	2007	0.4899
14	Petr Prucha (2006-2011, RW)	2006	0.4772
15	Alex Ovechkin (2006-2018, LW)	2015	0.4767
16	Alex Ovechkin (2006-2018, LW)	2013	0.4727
17	Thomas Vanek (2006-2018, LW)	2009	0.4718
18	Alex Ovechkin (2006-2018, LW)	2010	0.4718
19	Marian Gaborik (2001-2018, RW)	2007	0.4716
20	John LeClair (1991-2007, LW)	2003	0.4714

Table 2: Top 20 Players by Single-Season Lambda

	Season	Estimate	Actual	Error
1	2001	42.97	93	-50.03
2	2002	36.84	57	-20.16
3	2003	38.10	75	-36.90
4	2004	32.45	46	-13.55
5	2006	54.34	79	-24.66
6	2007	47.83	71	-23.17
7	2008	40.89	73	-32.11
8	2009	44.51	65	-20.49
9	2010	41.38	67	-25.62
10	2011	39.93	76	-36.07
11	2012	37.04	55	-17.96
12	2013	20.26	32	-11.74
13	2014	34.70	56	-21.30
14	2015	32.51	49	-16.49
15	2016	33.33	67	-33.67
16	2017	35.67	59	-23.33

Table 3: Forecasted Hat Tricks by Season (Season-by-Season Model)

	Season	Estimate	Actual	Error
1	2001	81.76	93	-11.24
2	2002	64.30	57	7.30
3	2003	70.15	75	-4.85
4	2004	54.86	46	8.86
5	2006	93.05	79	14.05
6	2007	83.89	71	12.89
7	2008	74.26	73	1.26
8	2009	77.08	65	12.08
9	2010	70.92	67	3.92
10	2011	67.37	76	-8.63
11	2012	67.45	55	12.45
12	2013	38.07	32	6.07
13	2014	61.89	56	5.89
14	2015	57.23	49	8.23
15	2016	58.79	67	-8.21
16	2017	60.77	59	1.77

Table 4: Forecasted Hat Tricks by Season (Player-by-Player Model)

	Season	Year_by_Year	Player_by_Player	Actual
1	2001	42.97	81.76	93
2	2002	36.84	64.30	57
3	2003	38.10	70.15	75
4	2004	32.45	54.86	46
5	2006	54.34	93.05	79
6	2007	47.83	83.89	71
7	2008	40.89	74.26	73
8	2009	44.51	77.08	65
9	2010	41.38	70.92	67
10	2011	39.93	67.37	76
11	2012	37.04	67.45	55
12	2013	20.26	38.07	32
13	2014	34.70	61.89	56
14	2015	32.51	57.23	49
15	2016	33.33	58.79	67
16	2017	35.67	60.77	59

Table 5: Forecasted Hat Tricks by Season (All Models)

	Season	Estimate	Actual	Error
1	2001	33.06	36	-2.94
2	2002	23.69	15	8.69
3	2003	26.68	29	-2.32
4	2004	19.00	15	4.00
5	2006	39.24	25	14.24
6	2007	36.11	19	17.11
7	2008	30.50	24	6.50
8	2009	33.55	23	10.55
9	2010	26.19	20	6.19
10	2011	27.77	22	5.77
11	2012	28.44	23	5.44
12	2014	23.20	18	5.20
13	2015	23.83	13	10.83
14	2016	21.79	26	-4.21
15	2017	23.90	23	0.90

Table 6: Second-half Forecasted Hat Tricks by Season

	Name	Estimate	Actual	Error
1	Alex Ovechkin (2006-2018, LW)	23.95	17	6.95
2	Jarome Iginla (1997-2017, RW)	14.18	12	2.18
3	Ilya Kovalchuk (2002-2013, LW)	13.82	13	0.82
4	Steven Stamkos (2009-2018, C)	12.15	8	4.15
5	Sidney Crosby (2006-2018, C)	11.81	10	1.81
6	Marian Hossa (1998-2017, RW)	11.51	7	4.51
7	Dany Heatley (2002-2015, LW)	10.53	9	1.53
8	Marian Gaborik (2001-2018, RW)	10.32	14	-3.68
9	Rick Nash (2003-2018, LW)	10.28	7	3.28
10	Jaromir Jagr (1991-2018, RW)	9.49	9	0.49
11	Evgeni Malkin (2007-2018, C)	9.39	11	-1.61
12	Patrick Marleau (1998-2018, C)	8.89	5	3.89
13	Corey Perry (2006-2018, RW)	8.43	9	-0.57
14	Vincent Lecavalier (1999-2016, C)	8.22	7	1.22
15	Teemu Selanne (1993-2014, RW)	7.83	6	1.83
16	Daniel Alfredsson (1996-2014, RW)	7.46	7	0.46
17	Milan Hejduk (1999-2013, RW)	7.46	5	2.46
18	Jeff Carter (2006-2018, C)	7.35	5	2.35
19	Martin St. Louis (1999-2015, RW)	7.30	8	-0.70
20	Thomas Vanek (2006-2018, LW)	7.24	9	-1.76
21	Eric Staal (2004-2018, C)	7.14	13	-5.86
22	Markus Naslund (1994-2009, LW)	6.88	8	-1.12
23	Zach Parise (2006-2018, LW)	6.87	5	1.87
24	Simon Gagne (2000-2015, LW)	6.39	3	3.39
25	Henrik Zetterberg (2003-2018, LW)	6.25	6	0.25
26	Patrick Kane (2008-2018, RW)	6.13	4	2.13
27	Daniel Sedin (2001-2018, LW)	6.11	6	0.11
28	Jason Spezza (2003-2018, C)	5.89	7	-1.11
29	Patrik Elias (1996-2016, LW)	5.85	8	-2.15
30	Phil Kessel (2007-2018, RW)	5.83	5	0.83
31	Joe Sakic (1989-2009, C)	5.79	6	-0.21
32	Joe Pavelski (2007-2018, C)	5.74	4	1.74
33	Alexander Semin (2004-2016, LW)	5.71	7	-1.29
34	Keith Tkachuk (1992-2010, LW)	5.54	1	4.54
35	Alex Kovalev (1993-2013, RW)	5.47	10	-4.53
36	Daniel Briere (1998-2015, C)	5.45	5	0.45
37	Bill Guerin (1992-2010, RW)	5.37	8	-2.63
38	Mats Sundin (1991-2009, C)	5.35	2	3.35
39	Jonathan Toews (2008-2018, C)	5.33	4	1.33
40	Olli Jokinen (1998-2015, C)	5.25	7	-1.75
41	Pavel Datsyuk (2002-2016, C)	5.17	0	5.17
42	Mike Cammalleri (2003-2018, C/W)	5.16	6	-0.84
43	John Tavares (2010-2018, C)	5.12	6	-0.88
44	Joe Thornton (1998-2018, C)	5.12	4	1.12
45	Shane Doan (1996-2017, RW)	4.97	2	2.97
46	Patrick Sharp (2003-2018, LW)	4.95	4	0.95
47	Ryan Smyth (1995-2014, LW)	4.95	3	1.95
48	Glen Murray (1992-2008, RW)	4.88	3	1.88
49	Max Pacioretty (2009-2018, LW)	4.88	6	-1.12
50	James Neal (2009-2018, LW)	4.86	6	-1.14
51	Brian Gionta (2002-2017, RW)	4.83	1	3.83
52	Pavel Bure (1992-2003, RW)	4.71	5	-0.29
53	Jonathan Cheechoo (2003-2010, RW)	4.67	9	-4.33
54	Jason Arnott (1994-2012, C)	4.60	6	-1.40
55	Brendan Shanahan (1988-2009, LW)	4.60	6	-1.40
56	Jamie Benn (2010-2018, LW)	4.27	2	2.27

Table 7: Top 56 Players by Forecasted Hat Tricks

	Name	Observed Hat Tricks	Top 56 Expected?
1	Alex Ovechkin (2006-2018, LW)	17	Yes
2	Marian Gaborik (2001-2018, RW)	14	Yes
3	Eric Staal (2004-2018, C)	13	Yes
4	Ilya Kovalchuk (2002-2013, LW)	13	Yes
5	Jarome Iginla (1997-2017, RW)	12	Yes
6	Evgeni Malkin (2007-2018, C)	11	Yes
7	Alex Kovalev (1993-2013, RW)	10	Yes
8	Sidney Crosby (2006-2018, C)	10	Yes
9	Corey Perry (2006-2018, RW)	9	Yes
10	Dany Heatley (2002-2015, LW)	9	Yes
11	Jaromir Jagr (1991-2018, RW)	9	Yes
12	Jonathan Cheechoo (2003-2010, RW)	9	Yes
13	Scott Hartnell (2001-2018, LW)	9	No
14	Thomas Vanek (2006-2018, LW)	9	Yes
15	Bill Guerin (1992-2010, RW)	8	Yes
16	Markus Naslund (1994-2009, LW)	8	Yes
17	Martin St. Louis (1999-2015, RW)	8	Yes
18	Patrik Elias (1996-2016, LW)	8	Yes
19	Steven Stamkos (2009-2018, C)	8	Yes
20	Alexander Semin (2004-2016, LW)	7	Yes
21	Daniel Alfredsson (1996-2014, RW)	7	Yes
22	Erik Cole (2002-2015, LW)	7	No
23	Jason Spezza (2003-2018, C)	7	Yes
24	Marian Hossa (1998-2017, RW)	7	Yes
25	Olli Jokinen (1998-2015, C)	7	Yes
26	Paul Kariya (1995-2010, LW)	7	No
27	Rick Nash (2003-2018, LW)	7	Yes
28	Steve Sullivan (1996-2013, RW)	7	No
29	Tyler Seguin (2011-2018, C)	7	No
30	Vincent Lecavalier (1999-2016, C)	7	Yes
31	Brendan Shanahan (1988-2009, LW)	6	Yes
32	Daniel Sedin (2001-2018, LW)	6	Yes
33	Drew Stafford (2007-2018, RW)	6	No
34	Henrik Zetterberg (2003-2018, LW)	6	Yes
35	James Neal (2009-2018, LW)	6	Yes
36	Jason Arnott (1994-2012, C)	6	Yes
37	Jason Blake (1999-2012, LW)	6	No
38	Joe Sakic (1989-2009, C)	6	Yes
39	John Tavares (2010-2018, C)	6	Yes
40	Martin Havlat (2001-2016, RW)	6	No
41	Max Pacioretty (2009-2018, LW)	6	Yes
42	Mike Cammalleri (2003-2018, C/W)	6	Yes
43	Radim Vrbata (2002-2018, RW)	6	No
44	Teemu Selanne (1993-2014, RW)	6	Yes
45	Brett Hull (1987-2006, RW)	5	No
46	Daniel Briere (1998-2015, C)	5	Yes
47	Jeff Carter (2006-2018, C)	5	Yes
48	Mark Parrish (1999-2011, RW)	5	No
49	Milan Hejduk (1999-2013, RW)	5	Yes
50	Patrick Marleau (1998-2018, C)	5	Yes
51	Pavel Bure (1992-2003, RW)	5	Yes
52	Peter Bondra (1991-2007, RW)	5	No
53	Phil Kessel (2007-2018, RW)	5	Yes
54	Steve Reinprecht (2000-2011, C)	5	No
55	Todd Bertuzzi (1996-2014, RW)	5	No
56	Zach Parise (2006-2018, LW)	5	Yes

Table 8: Top 56 Players by Observed Hat Tricks