Predicting Coin Flips: Using Resampling and Hierarchical Models to Help Untangle the NHL's Shootout

Michael Lopez* — Skidmore College (mlopez1@skidmore.edu)

Michael Schuckers — St. Lawrence University (schuckers@stlawu.edu)

February 6, 2016

Abstract

Roughly 14% of regular season National Hockey League games since the 2005-06 season have been decided by a shootout, and the resulting allocation of points has impacted playoff races each season. But despite interest from fans, players, and league officials, there is little in the way of published research on team or individual shootout performance. This manuscript attempts to fill that void. We present both generalized linear mixed model and Bayesian hierarchical model frameworks to model shootout outcomes, with results suggesting that there are (i) small but significant talent gaps between shooters, (ii) marginal differences in performance among netminders, and (iii) few, if any, predictors of player success after accounting for individual talent. We also provide a resampling strategy to highlight a selection bias with respect to shooter assignment, in which coaches choose their most skilled offensive players early in shootout rounds and are less likely to select players with poor past performances. Finally, given that per-shot data for shootouts does not currently exist in a single location for public use, we provide both our data and source code for other researchers interested in studying shootout outcomes.

Word count: 5,100 words

*: Corresponding author

Keywords: hockey; shootouts; hierarchical models; Bayesian models; reproducibility

1 Introduction

Following the locked out 2004-05 regular season, the National Hockey League instituted a shootout to determine winners of regular season games that finished overtime still tied. Shootouts in hockey take a similar form to penalty kicks in association football (soccer) matches. In the NHL's adaptation, both teams take alternating penalty shots three times. If the teams are still tied after those three rounds, then the teams complete single rounds until one team scores and the other does not.

To ensure that the shootout was taken seriously, the NHL changed its point system for the 2005-06 season, awarding teams the same number of points in the standings, two, as was awarded for a regulation or overtime win. Teams that lose in overtime or in the shootout are awarded one point, while teams that lose in regulation receive no points.

Roughly 1 in 7 NHL games have reached a shootout since it's implementation, which, in an 82-game season, equates to an average of 11 or 12 points a year for each team that are decided by the post-overtime competition. Given this point system, shootout performance can make or break an NHL team's season, as the resulting league standings often yield little separation between the top playoff qualifiers and teams not participating in the playoffs. For example, San Jose (No. 12 finisher, 89 points) finished within 15 points of Nashville (No. 3, 104 points) in the final 2014-15 standings in the 14-team Western Conference. Perhaps unsurprisingly, Seppa (2009) identified the shootout as a deciding factor in which teams make the playoffs. With NHL teams earning a disproportionate amount of the profit that comes in playoff ticket sales (Leeds & Von Allmen, 2004), the shootout also has financial ramifications.

While the league's revenue has increased since the shootout's implementation (Stubits, 2014), critics of the shootout have called it a 'gimmick' (Feschuk, 2014). Notably, Calgary Flames president Brian Burke called it a 'circus stunk...horse (bleep)' (Brough, 2014). Much of the frustration with respect to the shootout lies in its randomness, as most consider the outcome to be no different than a 'coin flip' (Feschuk, 2014; Schuckers, 2009). But is it? Despite the shootout's consistent and powerful influence on the NHL, there is little published work on the factors that drive individual and team performance. Moreover there is no centralized shot-by-shot data set for use; even the league's own website, `www.nhl.com`, computes only aggregated player statistics for the shootout. This stands in stark contrast to professional soccer, where the shootout has been studied extensively. See Pollard & Reep (1997), McGarry & Franks (2000), Jordet et al. (2007), Jordet (2009), Apesteguia & Palacios-Huerta (2010),Wood et al. (2015), to name a few, for examples of analyses for soccer.

This manuscript makes four major contributions. First, we look at overall team and individual shootout performances to identify how players perform relative to chance alone. Second, we use advanced regression modeling strategies to identify if players perform better or worse under pressure, as well as to quantify and compare within shooter and within goalie variability. Our results suggest that, while it is difficult to distinguish between the performance of most players, there are certain shooters and a few goalies who have significantly outperformed expectations. Meanwhile, performance does not appear to vary under pressure.

Third, we implement a set of simulations that identify how coaches select players for the shootout, as well as to identify the financial worth of the league's top players to their teams. We estimate that the league's best shooters and goalies have been worth more than half of a million American dollars per year to their teams on shootout performance alone. Finally, given that per-shot data is missing in easily analyzable format, as well as the increased awareness regarding the importance of reproducible scientific research (Peng, 2011), we provide all of our data and code, so that future researchers can confirm our results while performing their own analyses.

2 Hockey and Shootouts

There were 1583 shootouts in the ten full seasons between 2005-06 and 2014-15. Most analyses of these shootouts has come in the form of blog posts. Among these are three articles by Gabe Desjardins (writing as Hawerchuk). In the first of these, Desjardins looked at the distribution (in shots) of the shootout through March of 2009 and compared that to what would be expected given a Bernoulli model with success probability 33% for each shot (Desjardins, 2009a). Later, Desjardins wrote about comparing past performance in the shootout to future performance for shooters (Desjardins, 2009b) and for goalies (Desjardins, 2009c). His conclusion based upon these position specific analyses is that there is not much difference in the performance of the top players from the bottom players. Schuckers (2009) provided a simple Bayesian analysis of the first five years of the shootout in the NHL using a probit model that included a term for each shooter and each goalie, assuming a flat prior on each parameter. We build upon that analysis below. Schuckers' conclusion was that the shootout is randomness, a crapshoot. Eric Tulsky (aka Eric T.) (2009) used additional data and also concluded that there was not enough evidence to conclude that there were any players whose performance stood out. More recently, Schuckers & Nelson (2013) looked at a team's choice to shoot first or second, finding no significant difference.

There is scant peer-reviewed literature on the subject. McEwan et al. (2012) identified that visiting team players performed better in win-imminent situations (shooting to win) relative to home team players, but that shooters on home teams performed significantly better in situations when trying to avoid a loss. Jones (2013) looked at the impact of home advantage across a variety of sports, finding that in the first four years of the NHL shootout, there was no home advantage. In neither of these papers did the authors control for the individual talent levels among goaltenders or shooters. While mostly unrelated to our goals, Hurley (2005) took a statistical look at alternatives to the shootout, and Hurley (2008) analyzed approximately a season and a half of the NHL shootout to model its length.

In contrast to hockey, there is a sizable peer-reviewed literature on shootouts in soccer, which take the form of alternating penalty kicks. Pollard & Reep (1997) discuss shootouts as part of an in-depth statistical study of events in soccer, and the optimal order of players is analyzed by McGarry & Franks (2000). Jordet et al. (2007) used modeling to suggest that there is evidence that shooters choke in high pressure situations, the results of which were expanded upon by Jordet (2009). Apesteguia & Palacios-Huerta (2010) found that there is a significant difference between going first and second — first is significantly better. Additionally, Apesteguia & Palacios-Huerta (2010) showed that scoring rates increase in later rounds for teams that are leading and scoring rates decrease for teams that are behind. In fact, even team uniform color and 'pre-penalty gaze' have been discussed as predictors of performance (Greenlees et al., 2008). We study some of these identical covariates to consider their impact on the scoring rates in hockey shootouts.

3 Methods

3.1 Data Collection

We collected data on all shootouts using play-by-play game output provided by `www.nhl.com`. The data set contains 10839 shootout attempts from the 2005-06 through 2014-15 regular seasons, 3604 (33.3%) of which were successfully converted by the shooter. There are 164 different goalies and 828 different shooters in these data. The shooter taking the shot, his position (defense or forward), the goalie facing the shot, whether or not the shot resulted in a goal, the current score of the shootout, the home and away teams, and shot number (e.g. first, second, fifth) were all recorded.

We notate these variables as follows. Our outcome is $Goal_{ijk}$, whether or not shooter j scored against goalie k on attempt i, for $i = 1 \ldots 10839$, $j = 1, \ldots 828$, and $k = 1, \ldots 164$. Let $Defense_i$ be an indicator for whether or not the shooter of shot i is a defender, and let $Visiting_i$ be an indicator for whether or not shot i is attempted by the visiting team. Given the literature on how pressure can impact soccer penalty kicks, we define each shot as follows.

$$Status_i = \begin{cases} WinImminent, & \text{if the shooter ends the game with a goal on shot } i \\ LossImminent, & \text{if the shooter needs a goal on shot } i \text{ to keep game alive} \\ Early, & \text{otherwise.} \end{cases} \quad (1)$$

3.2 Team analysis

We first contrast win percentage for home versus away teams. Home teams have been given the opportunity to shoot or defend first in every season since 2005-06. In the 2005-06 season, home teams were required to shoot second. If there has been a home advantage, we expect the home team to win more than 50% of shootouts. A one-proportion z-test will be used to judge significance. Given evidence that shooting first or second matters in soccer (Apesteguia & Palacios-Huerta, 2010), we perform a similar analysis to judge if shootout order matters in hockey as far as team success. Finally, we will check how well a team's performance translates from one season to the next using a scatter plot, with the x-axis being the team's shootout win percentage in one year and the y-axis that team's win percentage in the following year.

3.3 Individual analysis

Judging shooter (or goalie) success by goal (save) percentage alone is difficult, given that many players have only participated in a few shootouts. For example, while there are 39 shooters who have scored on all of their attempts, none of these skaters have attempted more than two shots. As a result, it is difficult to contrast their perfect records to players having a substantially larger number of attempts.

In place of ranking players on goal or save percentages alone, funnel plots allow visualization of shooter and goalie performance while factoring in their relative sample sizes (Spiegelhalter, 2005). We briefly describe construction of a funnel plot for goalies; our plot for shooters is similarly constructed.

Our goalie funnel plot contrasts each goalie's career save percentage against his number of attempts. At the center of the y-axis lies the the population save rate, $\lambda = 0.668$, which, under a hypothesis that all goalies are identical, represents the expected value of each goalie's individual rate. Two confidence limits are shown, representing the bounds for individual performances lying two (95% limits) or three (99.8%) standard deviations above or below λ. As in Spiegelhalter (2005), we use the exact limits from the inverse Binomial distribution, which are preferred over a normal approximation with smaller sample sizes.

If shootouts are a random outcome with save probability λ, we expect roughly 95% and 99.8% of goalies to fall within two and three standard deviations of λ. If there are more players than we expect beyond these boundaries, it would provide evidence that there is some inherent skill in saving shootout attempts. Thus, we check for the relative number of goalies falling beyond these boundaries in the funnel plots as one measure of the shootout's true randomness. For shooters, funnel plots are centered at $1 - \lambda$.

3.3.1 Shooter selection bias

The results of McEwan et al. (2012) suggest that shooter performance may vary by the shootout's pressure. However, a critical component of the shootout is that while goalies almost universally remain in net for the duration of the shootout, coaches can place a shooter in any round until each of a team's shooters have one attempt apiece. As a result, there is the potential that if the league's best shooters are always shooting first, the association between a shot's pressure and decreased shooter performance could be accounted for by changes in shooter talent.

Table 1 shows the shootout rounds of the league's five best shooters, as will be identified in Section 4.2.1. More than 92% of these shooters' attempts have been in rounds one or two, including 80 attempts from Jonathan Toews, who has never taken a shot after round two.

Table 1: Round deployment among top shooters

Shooter	Rd. 1	Rd. 2	Rd. 3	Rd. 4+
Nielsen	48	23	0	1
Toews	71	9	0	0
Oshie	41	9	7	2
Kozlov	36	32	9	3
Christensen	47	2	2	3
Total	243	75	17	9

Given that shootouts can end as early as round 2 (two goals from one team and blanks from its opponent), it is understandable why coaches would deploy their best shooters early. However, such an allocation makes it difficult to look at player performance under pressure, given that the types of shooters that shoot early (with generally less pressure) are likely more talented than those that shoot later (generally more pressure). Perhaps a preferable question would be to contrast shooter performance under pressure among shooters who have taken a sufficient number of attempts in all rounds and under all pressures; this strategy is implemented in Section 3.4.

3.4 Models of NHL shootout outcomes

To simultaneously account for several variables that may be concurrently impacting shootout rates, as well as the dependency within individual shooter and goalie outcomes, we propose multiple modeling frameworks.

3.4.1 Hierarchical mixed models

To account for the selection bias in the allocation of the best shooters to earlier shootout rounds, we will use two samples from our data. The first, the full sample, will contain all shots, while the second, identified as the reduced sample, will include only the attempts by shooters with at least five attempts in each of the first three shootout rounds. Let $n_j = 828$ and $n_k = 164$ and $n_{j'} = 88$ and $n_{k'} = 156$ be the number of shooters and goalies in the full and reduced samples, respectively.

Let $\boldsymbol{X_i}$ be a design matrix including fixed effects for $Status_i$ ($WinImminent$ and $LossImminent$ attempts compared to no pressure ones), $Visiting_i$, and $Defense_i$. To account for the correlation in $Goal_{ijk}$ between attempts made by the same shooter or against the same goalie, let δ_j, $\delta_{j'}$, δ_k, and $\delta_{k'}$ be shooter and goalie specific random effects. These correspond to each shooter and each goalie in the full and reduced samples, respectively.

We propose three generalized linear mixed models, or subsets thereof, where, letting $\text{logit}(p) = \log(\frac{p}{1-p})$,

$$\text{logit}(P(Goal_{ijk} = 1)) = \alpha + \beta_T \boldsymbol{X_i} \tag{2}$$

$$\text{logit}(P(Goal_{ijk} = 1)) = \alpha + \beta_T \boldsymbol{X_i} + \delta_j + \delta_k \tag{3}$$

$$\text{logit}(P(Goal_{ijk} = 1)) = \alpha + \beta_T \boldsymbol{X_i} + \delta_{j'} + \delta_{k'} \tag{4}$$

where $\delta_j \sim N(0, \tau_j^2)$, $\delta_k \sim N(0, \tau_k^2)$, $\delta_{j'} \sim N(0, \tau_{j'}^2)$ and $\delta_k \sim N(0, \tau_{k'}^2)$.

Model (2) features a set of fixed effects but no random effects, while (3) includes random effects for both shooter and goalie. Model (4) uses only players in the reduced set of shots. If there is large variability between shooter performances after accounting for $\mathbf{X}$, we would expect large τ estimates. We can also use the estimated τ terms to contrast the relative amount of variability between player performances.

3.4.2 Bayesian Model of NHL Shootout outcomes

Similar to the work of the *GLM* model above, we also fit a hierarchical Bayesian model to assess the shootout. Our model followed a similar structure to that proposed by Albert & Chib (1993), where

$$
\begin{aligned}
Goal_{ijk} &\sim Bernoulli(z_{ijk}), \\
z_{ijk} &\sim TruncMVN(\boldsymbol{\theta}^T \mathbf{X_i}, \mathbf{1}), \\
\boldsymbol{\theta}^T = (\alpha, \boldsymbol{\delta}_g, \boldsymbol{\delta}_s, \beta)^T &\sim MVN(0, \boldsymbol{\tau}^T \mathbf{I}), \\
\tau_\alpha^2 &\sim \text{Inverse } \chi^2(5), \\
\tau_g^2 &\sim \text{Inverse } \chi^2(5), \\
\tau_s^2 &\sim \text{Inverse } \chi^2(5), \\
\tau_\beta^2 &\sim \text{Inverse } \chi^2(5),
\end{aligned}
\qquad (5)
$$

such that $\boldsymbol{\tau}$ is a concatenation of the precision parameters with $\boldsymbol{\tau} = (\tau_\alpha^2, \tau_j^2 \mathbf{1}_j, \tau_k^2 \mathbf{1}_k, \tau_\beta^2 \mathbf{1}_4)^T$ where $\mathbf{1}_t$ is a vector of ones of length t. Then τ_α is the precision for the intercept term; τ_g is the precision for the goalie terms; τ_s is the precision for the shooter terms; and τ_β is the precision for the other covariates in our model. We are thus assuming that there is a latent variable for each shot, z_{ijk}, which depends on j, k, $Status_i$, $Visiting_i$, and $Defense_i$. The TruncMVN represents a truncated multivariate Normal distribution which is truncated to only positive values if $Goal_{ijk} = 1$ and only negative values if $Goal_{ijk} = 0$. The expected value of the latent variable, z_{ijk}, is determined by the product of the model coefficients, $\boldsymbol{\theta}$, and the design matrix $\mathbf{X}$, where $\mathbf{X}$ is the matrix of covariates with each row corresponding to $\mathbf{X_i}$, the observed covariates for the i^{th} shootout shot attempt. The precision of the z_{ijk}'s is unity, again following Albert & Chib (1993).

The model coefficients, denoted by θ, are a concatenation of the parameters for goalies, shooters and the other covariates in the model, namely *Status*, *Visiting*, and *Defense*. The prior for these coefficients, $\boldsymbol{\theta}$, is a multivariate Normal distribution (MVN) with mean zero and precision matrix a vector $\boldsymbol{\tau}$ times the identity matrix. We follow Gelman (2006) in using an Inverse χ^2 for the prior distribution of our precision parameters, the τ's. Our choice of five degrees of freedom for the prior distribution of the τ's is one that assumes some stability to the model precision but that can be "swamped" by the large number of observations here.

Consent for the data collection and analysis was given by the Institutional Review Board at Skidmore College. Statistical analysis was conducted using the *R* statistical software (R Core Team, 2015), and Models (2) through (4) were fit using the 'lme4' package (Bates et al. (2015)). All *R* code and data can be found at this url: `https://github.com/NHLshootouts/NHL-shootouts`.

4 Results

4.1 Team level results

The road team won 52.2% of the NHL's 1583 shootouts, a percentage that is not significantly different than 0.5 ($p - value = 0.09$). Interestingly, road team success peaked in the 2010-11 season. Figure 1 shows the yearly win percentage for road teams in shootouts, as well as 95% confidence intervals for each yearly sample proportion. In all but the 2010-11 season, during which road teams won 91 of the 149 shootouts (61.0%), the road team win percentage has been within ± 5 points of 50%.

*** ***

Figure 1: Road team yearly win percentage in the shootout, with 95% confidence limits

All together, 49.6% of teams shooting first eventually won the shootout, a percentage that did not noticeably vary by home or away team. Despite the decision to go first making no apparent difference on win likelihood, an overwhelming 77.6% of teams chose to go first.

Figure 2 shows a scatter plot of team shootout win percentage in one year versus its win percentage in the year prior, along with a locally weighted smoothed line and its 95% confidence limits.

There is a slight positive increase to the line, indicating that teams winning a larger fraction of shootouts in one year may be more likely to win more shootouts in the next year. However, at all win percentages, the deviations from 50% are not significant, indicating that, by and large, team-level shootout ability does not correlate from one season to the next.

*** ***

Figure 2: Year to Year team winning percentage in the shootout, along with a locally weighted smoother and its 95% confidence limits

4.2 Individual level results

4.2.1 Descriptive statistics and visualizations

Shooters in low pressure situations successfully converted 33.9% of attempts, versus 32.5% of win imminent and 30.5% of loss imminent attempts. As judged by using a χ^2-test, these differences were significant (p-value = 0.03). Defensemen were only successful on 27.9% of attempts, compared to 33.6% for forwards (p-value = 0.003). However, coaches used defensemen on only 662 (6.1%) opportunities.

Figures 3 and 4 show funnel plots for skaters (all) and goalies (at least 10 attempts), respectively, along with confidence limits for individual performances that lie two or three standard deviations above or below the league-wide averages.

*** ***

Figure 3: Goalie attempts and save rates, with dotted lines representing 95% and 99.8% confidence limits for individual goalie performance

In both of Figures 3 and 4, the majority of players are clustered within the confidence limits. However, particularly with shooters in Figure 4, there are a handful of players who are outperforming expectations, lying more than two or three standard deviations above the league average. Among the 15 shooters with at least 75 shootout attempts, 6 (40%) lie above the 97.5^{th} percentile, and 2

(13%) lie above the 99.8^{th} percentile, at least with respect to our expectations under league-wide randomness. Of the 11 goalies with at least 200 attempts, 3 (27%) lie above the 97.5^{th} percentile. This suggests, particularly among shooters, that there is variability in performance that exceeds what we would expect due to chance alone.

*** ***

Figure 4: Shooter attempts and goal rates, with dotted lines representing 95% and 99.8% confidence limits for individual shooter performance

Relative to league-wide expectations, as well as their sample sizes, the league's best shooters are Eric Christensen (54 attempts, 53.7% success rate), Franz Nielsen (72, 52.8%), TJ Oshie (59, 52.5%), Victor Kozlov (80, 53.8%), and Jonathan Toews (80, 50%). Using a similar strategy, Henrik Lundvquist (311, 73.5% save percentage) rates as the league's most impressive netminder in shootouts, while Nicklas Backstrom (185, 54.1%) rates as its worst.

For interactive plots that allow scroll-over of each player, visit `https://plot.ly/~mlopez1/23.embed` (goalies) and `https://plot.ly/~mlopez1/17.embed` (shooters).

4.2.2 Regression model results

Following the advice of Gelman et al. (2002), and in place of tables with coefficient estimates, we use whisker plots of the estimated coefficients from Models (2)-(4), along with their 95% confidence intervals, as shown in Figure 5. Coefficients are presented on the log-odds scale, such that the positive estimates are associated with increased goal likelihood. (These estimates are available upon request, and also appear in the *R* code output.)

*** ***

Figure 5: Coefficient estimates from mixed models of shootout outcomes, with 95% confidence intervals

In Model (2), there is strong evidence that shooters perform significantly worse in loss imminent situations (Odds Ratio (OR) 0.85, 95% CI 0.74-0.96). This is shown in Figure 5 by an error bar for ‘Loss Imminent‘ situations that does not include 0. However, after accounting for the individual talent of goalies and shooters, as in (3) and (4), there is limited to no evidence that shooters are any better or worse under pressure. In the model fit on the reduced subset of the data, estimates for shots taken under both pressures are near 0.

Using the output from Model (2), the odds of a goal on shots by a member of the visiting team are about 8% higher (OR 1.08, 95% CI 1.00-1.18) than for home team attempts. Coefficient estimates for $Visiting$ are robust to model specification.

The estimated random effect terms from Models (3) and (4) are $\hat{\tau}_j = 0.191$, $\hat{\tau}_k = 0.178$, $\hat{\tau}_{j'} = 0.104$, and $\hat{\tau}_{k'} = 0.073$. This suggests that the variability in shootout skill between shooters (j) is slightly larger than that between goalies (k). Further, the overall magnitudes of these estimates are far from zero, suggesting non-random variability applicable to both positions. Using the estimated shooter and goalie intercepts from Model (3), Victor Kozlov, Jonathan Toews, Frans Nielsen, and TJ Oshie rank as the league's best shooters, with Michael Ryder, Tomas Plekanec, Clarke MacArthur, and Martin Havlat ranking as its worst. Marc-Andre Fluery and Henrik Lundqvist rank as the league's best goalies (lowest random effects), with Nicklas Backstrom and Martin Biron ranking as the the worst.

By and large, the results of the Bayesian approach are similar to those found by using frequentist methods. Using the model given in the previous section, we ran a single MCMC with a burn-in of 1000 iterations followed by 10000 iterations where we thinned the chain by taking every 10th iteration. Thus we have 1000 draws from the joint posterior distribution of the model parameters. Each of the posterior 99% credible intervals for the covariates in the Bayesian model (5), $Defense$, $Visiting$, and $Status$, includes 0. This suggests that the *a posteriori* probability of these variables being associated with shootout outcomes is small. The posterior mean for the standard deviation of shooters ($\hat{\tau}_j^{-1} = 0.222$) is higher than that of goalies ($\hat{\tau}_k^{-1} = 0.127$). Although these scales are different than in the mixed models in Model (3) and Model (4), there is agreement that there is a larger variation among shooters than among goalies. The Bayesian model does not yield any goalies that have 99% credible intervals that do not include zero. Petteri Nummelin, Jakob Silfverberg, Victor Kozlov, Frans Nielsen, and TJ Oshie rank as the top five shooters in the posterior summaries, as judged by using the lower bound of each player's credible interval among players whose intervals do not include zero.

5 Simulations

5.1 Resampling the shootout under randomness

Given that coaches are responsible for which shooters get the most opportunities, it is reasonable to expect that teams use past shooter performance in order to dictate strategy. Under this hypothesis, the best shooters would continue to get more attempts while shooters that struggle, either initially or eventually, would be passed over. We test this hypothesis as follows.

Let $SH_j(x)$ be the cumulative shooting percentage of shooter j after shot x, such that

$$SH_j(x) = \frac{\sum_{i=1}^{x} Goal_{ij}}{x}, \tag{6}$$

where $Goal_{ij}$ is an indicator for whether or not shooter j scored on shot i. Our interest lies in comparing how well the observed pattern of shooter deployment fits the expectations given a more random assignment of shooters to attempts.

To identify what $SH_j(x)$ would look like if gaps in shooter talent remained as wide as currently estimated but shooter allocation was independent of past success, we simulate under the following conditions. First, the overall league mean of each skater's goal percentage is taken to be the overall mean (33.25%, log-odds -0.695). Second, using $\hat{\tau}_j$, we assume that each shooters' random intercept comes from the $N(-0.695, \hat{\tau}_j = 0.191)$ distribution. We simulate intercepts for each shooter, transforming to get simulated probabilities. Each actual shooter (with known n_j total attempts) is assigned a simulated probability.

To compare the observed and simulated $SH_j(x)$, we use spaghetti plots (Figure 6). The left panel of Figure 6 shows observed $SH_j(x)$ for all j, while the right panel shows one example of a simulated $SH_j(x)$ for all j. In each panel, the grey line reflects each player's cumulative shootout percentage over time, while the black dot reflects the players eventual percentage after all attempts are complete. Points are jittered to account for overlapping probabilities.

There are several differences between the observed and simulated tracks. In the simulated

*** ***

Figure 6: Observed and simulated shootout percentage tracks, given identical sample size of attempts

panel, far more players with high career percentages do not have additional opportunities. Related, in the simulated tracks panel, more players with a relatively poor performance are given additional opportunities. Meanwhile, the larger number of black dots in the bottom left of the observed panel suggests that in practice, most of these relatively poor shooters are no longer awarded opportunities.

Finally, the best players in the observed tracks panel have higher goal rates than in the simulated tracks. Combined with the random effect plots from Section 4, this suggests a possible skewed right distribution of the shooter-specific random intercepts. Most shooters converge around the overall mean of 33%; however, the poor ones are no longer given opportunities. Relative to what we would expect due to chance, the best shooters continue to outperform expectations.

5.2 Shooter and goalie value added

We estimate the relative importance of shooters and goalies with respect to team success in the shootout under two assumptions, $A1$ and $A2$. Under $A1$, we estimate the net impact of adding one of the league's best shootout players (goal percentage, 51.7%) to a league average team, in terms of additional points gained per season towards seasonal standings. Under $A2$, we estimate the net impact of the league's best goalie (goal percentage, 26.3%) playing for a team comprised of league-average shooters.

Against league average teams, teams under $A1$ and $A2$ would win roughly 56% and 60% of shootouts, respectively. The minimum number of yearly shootouts a team played between the 2005-06 and 2014-15 seasons was 6, the mean 11.2, median 12, and the maximum was 21. Assuming that the number of games in a season that teams reach the shootout follows a Poisson distribution with parameter 11.2, we simulated 1000 team-seasons, comparing $A1$ and $A2$ to the performance of league average teams in terms of expected points added. Implicit in these simulations is that there is no association between a teams' shootout ability and the frequency of shootouts they reach in a given season. This is reasonable; since the 2005-06 season, the correlation between a teams yearly shootout win percentage and its number of total shootouts is essentially zero (0.03).

Table 2 shows the mean, median, 2.5^{th}, and 97.5^{th} quantiles of the simulated points added under $A1$ - $A2$. Teams with the league's best shooter expect an additional 0.67 points towards the seasonal standings, on average, although in 47% of simulated seasons such a team would do worse than or the same as a league-average shootout team. Under $A2$, teams with the top goalie and league-average shooters do better than a league-average shootout team in 60% of seasons, picking up an average of 1.14 points per season.

Table 2: Net points towards season standings added

Assumption	Mean points added	Median (2.5^{th}-97.5^{th} percentiles)
$A1$	0.67	1 (-4, 5)
$A2$	1.14	1 (-4, 6)

Using 1000 simulated seasons

Given that the relative value of an additional point in the standings for an NHL team has been valued at roughly a million American dollars (Patrick, 2014), it is reasonable to argue that, on shootout performance alone, Nielsen, Toews, and Oshie have been worth roughly $670,000 in expectation, with Lundvquist and Fleury worth just over a million.

Incidentally, the expected shootout win percentages for teams with top shooters or goalies match empirical evidence. The New York Islanders, Chicago Blackhawks, and the St. Louis Blues have won 54.5% of shootouts since the 2007-2008 season, the first in which shootout stars Nielsen, Toews, and Oshie joined their respective teams. The Pittsburgh Penguins and the New York Rangers, featuring the league's best goalies in Fleury and Lundvquist since the 2005-06 season, have won 61.4% of shootouts over this time span.

6 Discussion & Conclusion

In contrast to much of the current literature, we present evidence that NHL shootouts are not an entirely random outcome, with the non-randomness coming in a couple of subtle forms.

First, although current literature on player skill in the shootout has argued that shootout results are what one would expect due to chance alone, we identify that there is significant non-random variability in player performance, at least among shooters. This is identified both using funnel plots and in regression modeling. Interestingly, while there is a larger variation in talent among shooters than goalies, it is more difficult for shooters to impact team performance as they, in all likelihood, only shoot once per contest. Given their performances over the past decade on shootouts alone, we find that shooters have been worth about two-thirds of a point per season, and goalies worth about a point per season, in expectation, although in many seasons it is difficult to distinguish these differences from random variation.

Second, using simulations, we identify a selection bias with respect to shooter allocation; those with past success are repeatedly given opportunities, while those who miss on successive attempts are less likely to be chosen again. Related, we find that the best shooters have been used by their coaches in the first and second rounds of the shootout.

In addition to looking at claims of randomness, our regression model results also give information regarding correlates of player success. Although McEwan et al. (2012) identified that player performance varies under pressure, we find no such evidence after accounting for individual shooter and goalie talent. One plausible explanation for this difference is the round deployment of the best shooters, with the best ones going early in shootouts, which could unnecessarily link poor performance from the less talented shooters with performance under pressure.

All together, evidence suggests that while there are few, if any, predictors of shootout success, there is a moderate amount of within-shooter and a small amount of within-goalie variability. Given the league's recent resistance to the shootout, noteworthy changes were made during the summer prior to the 2015-16 season. In place of a five minute overtime with each team using four skaters apiece, teams are now playing with three skaters apiece. In anticipation of such a system, Pettigrew (2015) used historical scoring rates to estimate that the fraction of overtimes reaching a shootout would drop from roughly 60% to 43% with the implemented change. Assuming teams play overtime at the same frequency as in past seasons, we would expect between three and four fewer shootouts per team in each season. For team officials, this would result in a corresponding drop to the valuation of player performances mentioned earlier. Of course, if the frequency of overtime games continues to increase, as it has since the 2004-05 lockout (Lopez, 2013), the drop off may not be as severe.

While research in professional soccer has identified evidence of players choking under pressure (Jordet, 2009), no such evidence is found in hockey. One hypothesis for this difference is that in soccer, shooters control most of the shootout's outcome, as by and large, they can score with accurate kicks. In hockey, meanwhile, goalies control as much of the outcome, if not more, than the shooter. Relatedly, the overall fraction of goals in soccer lies around 80%, relative to the lower 33% goal rate in hockey. So, whereas more pressure may fall on shooters in soccer, perhaps shooters and goalies feel less pressure to succeed on any given attempt in hockey.

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*, 669-679.

Apesteguia, J., & Palacios-Huerta, I. (2010, December). Psychological pressure in competitive environments: Evidence from a randomized natural experiment. *American Economic Review*, *100*, 2548-2564.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4 [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lme4` (R package version 1.1-8)

Brough, J. (2014). *Burke calls the shootout a circus stunt.* `http://nhl.nbcsports.com/2014/02/28/burke-calls-the-shootout-a-circus-stunt/`. (Accessed December 31, 2015)

Desjardins, G. (2009a, March). *Shootout length: Model vs actual.* `http://behindthenet.ca/blog/2009/03/shootout-length-model-vs-actual.html`. (Accessed September 4, 2015)

Desjardins, G. (2009b, October). *Shootouts: Does past performance mean anything?* `http://www.arcticicehockey.com/2009/10/19/1089029/shootouts-does-past-performance`. (Accessed September 4, 2015)

Desjardins, G. (2009c, October). *Shootouts: Goaltender true talent.* `http://www.arcticicehockey.com/2009/10/20/1089388/shootouts-goaltender-true-talent`. (Accessed September 4, 2015)

Eric Tulsky (aka Eric T.) . (2009). *Is the shootout really a crapshoot?* `http://www.broadstreethockey.com/2012/2/13/2789709/yes-it-is`. (Accessed June 15, 2015)

Feschuk, D. (2014, March). *No end in sight to nhls gimmick.* `http://www.thestar.com/sports/hockey/2014/03/12/no_end_in_sight_to_nhls_gameending_gimmick_the_shootout_feschuk.html`. (Accessed September 4, 2015)

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Gelman, A., Pasarica, C., & Dodhia, R. (2002). Let's practice what we preach: turning tables into graphs. *The American Statistician*, *56*(2), 121–130.

Greenlees, I., Leyland, A., Thelwell, R., & Filby, W. (2008). Soccer penalty takers' uniform colour and pre-penalty kick gaze affect the impressions formed of them by opposing goalkeepers. *Journal of sports sciences*, *26*(6), 569–576.

Hurley, W. J. (2005). Overtime or shootout: Deciding ties in hockey. In J. Albert, J. Bennett, & J. J. Cochran (Eds.), *Anthology of statistics in sports.* SIAM, Philadelphia, ASA, Alexandria, VA: ASA-SIAM Series on Statistics and Applied Probability, SIAM.

Hurley, W. J. (2008). On the length of nhl shootouts. *Journal of Modern Applied Statistical Methods*, *7*. `http://digitalcommons.wayne.edu/jmasm/vol7/iss1/27`.

Jones, M. (2013, January). The home advantage in individual sports: An augmented review. *Psychology of Sport and Exercise*, *14*, 397-404.

Jordet, G. (2009). Why do english players fail in soccer penalty shootouts? a study of team status, self-regulation, and choking under pressure. *Journal of sports sciences*, *27*(2), 97–106.

Jordet, G., Hartman, E., Visscher, C., & Lemmink, K. A. P. M. (2007). Kicks from the penalty mark in soccer: The roles of stress,skill, and fatigue for kick outcomes. *Journal of Sports Sciences*, *25*(2), 121-129. Retrieved from `http://dx.doi.org/10.1080/02640410600624020` doi: 10.1080/02640410600624020

Leeds, M., & Von Allmen, P. (2004). The economics of sports. *New York*.

Lopez, M. J. (2013). Inefficiencies in the national hockey league points system and the teams that take advantage. *Journal of Sports Economics*, 1527002513486654.

McEwan, D., Ginis, K. A. M., & Bray, S. R. (2012). "With the game on his stick": the home (dis) advantage in national hockey league shootouts. *Psychology of Sport and Exercise*, *13*(5), 578–581.

McGarry, T., & Franks, I. M. (2000). On winning the penalty shoot-out in soccer. *Journal of Sports Sciences*, *18*(6), 401-409. Retrieved from http://dx.doi.org/10.1080/02640410050074331 doi: 10.1080/02640410050074331

Patrick, D. (2014). *How much does a standings point cost in the nhl.* http://www.fearthefin.com/2014/8/22/6055583/how-much-does-a-standings-point-cost-in-the-nhl. (Accessed December 31, 2015)

Peng, R. D. (2011). Reproducible research in computational science. *Science (New York, Ny)*, *334*(6060), 1226.

Pettigrew, S. (2015). *How the 3 on 3 overtime rules would cut down nhl shootouts.* http://regressing.deadspin.com/how-those-3-on-3-overtime-rules-would-cut-down-nhl-shoo-1692120143. (Accessed December 31, 2015)

Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *46*(4), 541–550. Retrieved from http://dx.doi.org/10.1111/1467-9884.00108 doi: 10.1111/1467-9884.00108

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Schuckers, M. (2009). *NHL shootout is a crapshoot.* http://www.arcticicehockey.com/2010/12/29/1901842/nhl-shootout-is-a-crapshoot. (Accessed June 15, 2015)

Schuckers, M., & Nelson, Z. (2013, January). *Impact of going first on winning an nhl shootout.* http://statsportsconsulting.com/main/wp-content/uploads/ShootoutFirst.pdf.

Seppa, T. (2009, October). *Inside the vukota projections: Shootouts.*

Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance. *Statistics in medicine*, *24*(8), 1185–1202.

Stubits, B. (2014, June). *Nhl revenue to hit* $3.7b; cap likely to exceed 70$ *million.* http://www.cbssports.com/nhl/eye-on-hockey/24584103/report-nhl-revenue-to-hit-37-billion-cap-likely-to-exceed-70-million.

Wood, G., Jordet, G., & Wilson, M. R. (2015). On winning the "lottery": psychological preparation for football penalty shoot-outs. *Journal of sports sciences*(ahead-of-print), 1–8.