A Bayesian Adjusted Plus-Minus Analysis for the Esport Dota 2

Abstract

Analytics and professional sports have become linked over the past several years, but little attention has been paid to the growing field of esports within the sports analytics community. We seek to apply an Adjusted Plus Minus (APM) model, an accepted analytic approach used in traditional sports like hockey and basketball, to one particular esports game: Defense of the Ancients 2 (Dota 2). As with traditional sports, we show how APM metrics developed with Bayesian hierarchical regression can be used to quantify individual player contributions to their teams and, ultimately, use this player-level information to predict game outcomes. In particular, we first provide evidence that gold can be used as a continuous proxy for wins to evaluate a team's performance, and then use a Bayesian APM model to estimate how players contribute to their team's gold differential. We demonstrate that this APM model outperforms models based on common team-level statistics (often referred to as "box score statistics"). Beyond the specifics of our modeling approach, this paper serves as an example of the potential utility of applying analytical methodologies from traditional sports analytics to esports.

Keywords: esports, Bayesian Hierarchical Regression, Adjusted Plus Minus, Dota2

1 Introduction

Over the last several decades analytics have become an accepted and important part of sports. Along with individual teams using analytic methodologies to improve performance, a field of academic sports analytics has also emerged with many universities offering both Bachelor's and Master's Degrees in the field (Fisher, 2017). Simultaneously, electronic sports or esports have seen tremendous growth in popularity, creating an extensive fanbase (Kaytoue, Silva, Cerf, Meira Jr, and Raïssi, 2012). While on the surface there appears to be little overlap between the athletes who dominate professional sports and players who dominate video games such as Overwatch or League of Legends, both traditional sports and esports can benefit from similar analytic approaches.

In this manuscript, we demonstrate the utility of applying established sports analytics to esports using Defense of the Ancients 2 (Dota 2) as our game of interest. Just as sports analytics offer insight into top athletes, we demonstrate in this manuscript how analytics also offer insight into top esports players. To this end, we must first define a logical metric for scoring in the game. Unlike traditional sports, many esports lack a natural scoring mechanism: wins and losses are achieved through task accomplishment rather than obtaining points. To overcome this obstacle, we demonstrate that there are natural score-like proxies that help to quantify not only whether a team wins or losses but also how dominant one team is over another. Additionally, we use this score-like metric to quantify players' comparative strengths on offense and defense. Finally, we demonstrate how adjusting for competition provides a sharp increase in predictive power in esports when compared to simple box score metrics.

Dota 2 is an example of a Multiplayer Online Battle Arena (MOBA) similar to League of Legends and Call of Duty. The game was developed by Valve in 2011 and was conceived from player generated modifications to World of Warcraft (Drachen, Yancey, Maguire, Chu, Wang, Mahlmann, Schubert, and Klabajan, 2014). Dota 2 is a five on five game where each player chooses to play as one of over 100 different mythological characters, called heroes, each with their own strengths and weaknesses. The players select characters via a draft where teams alternatively ban and pick characters. The games are played in real-time and teams often communicate over headsets to come up with strategies for victory. The two teams, designated Dire and Radiant, compete on an unchanging virtual battlefield with the goal of destroying their opponents' base. The team that successfully destroys the opposing team's base is declared the winner. During the games, players can acquire gold by killing opposing heroes, destroying towers, and several other actions that move their team toward victory. Players can use gold to buy items or revive their hero (dota2.gamepedia.com, 2019) to provide additional strategic advantages.

Introduced in 2011, Dota 2 remains extremely popular. As of January 2019, there were still half a million unique players playing the game (steamcommunity.com, 2019). The game has spawned professional leagues who hold multiple tournaments each year, culminating in a grand championship tournament called The International. In 2018, The International (DOTA2.com, 2019) had a prize pool of over 25 million USD with a payout of over 11 million USD to the winning squad, a comparable payout to the major championships of traditional sports. For example, the 2018 PGA Championship had a total purse of 11 million USD (PGA.com, 2019) and the US Open tennis tournament had a total purse of 53 million USD with 3.8 million USD each awarded to the men's and women's singles champions (Marshall, 2019).

While the popularity of games such as Dota 2 has dramatically increased over the last several years, there has been relatively little focus on creating advanced metrics for ranking players in esports, especially when compared to the thriving field of sports analytics. The existing scholarly work focuses on applying machine learning techniques to predict game outcomes based on which characters are drafted, e.g. (Semenov, Romov, Korolev, Yashkov, and Neklyudov, 2016), or classifying player roles using supervised machine learning (Eggert, Herrlich, Smeddinck, and Malaka, 2015). These approaches, however, do not attempt to characterize individual players' contributions to the game; a nearly ubiquitous practice in traditional sports analytics.

In this manuscript, we present a Bayesian adjusted plus-minus metric for evaluating individual player contributions in Dota 2. This work extends regression-based adjusted plus-minus metrics from traditional sports such as professional hockey (Macdonald, 2011), (Macdonald, 2012), (Thomas, Ventura, Jensen, and Ma, 2013), (Schuckers and Curro, 2013), (Gramacy, Jensen, and Taddy, 2013), basketball (Rosenbaum, 2004), (Ilardi, 2007), (Ilardi and Barzilai, 2008, Sill, 2010, Okamoto, 2011), and soccer (Matano, Richardson, Pospisil, Eubanks, and Qin, 2018). Since Dota 2 and other esports have many similarities to traditional sports, we draw analogies to these familiar sports whenever possible.

Using a subset of professional Dota 2 games, we first demonstrate how a team's gold earned per minute can be used as a continuous proxy for win/loss, giving a natural way to score a Dota 2 match in a more granular way than simple classification. We then create a Bayesian adjusted plus-minus regression model that allows us to quantify both the offensive and defensive contribution of an individual Dota 2 player (in terms of gold) while accounting for the ability of a player's teammates and opponents. We show that this new metric outperforms player statistics that are analogous to box score statistics in traditional sports, and show that on a small set of out of sample games we can accurately predict the outcomes of 60% of games using adjusted plus-minus statistics, compared to a rate of 57% when using box score statistics. Furthermore on this set of games, the adjusted plus-minus approach is significantly better at detecting expected blow-outs that occur when one team is expected to perform much better than another. When the expected difference in gold is over 200 (a blow out), the APM model correctly predicts 70% of games whereas the box scores only correctly classify 60%. Finally, over a much larger sample of more recent games, we show how the APM model correctly classifies 66% of the games suggesting that the APM model is quite stable over time.

2 Why esports as a sport?

While we demonstrate that the analytical methods used in sports applies to esports, the similarities between the disciplines extend well beyond this. Similar to basketball, Dota 2, for example, is a five on five team game where each player has an individual skill set and the team may or may not be equal to the sum of the parts, depending on if the players have overlapping skill sets or complementary skill sets. Further, esports have players who have individual stats that may or may not be influenced by who their teammates are and what role they are playing, again similar to many traditional sports such as hockey, soccer, basketball, or football. In Dota 2, each of the heroes has different attributes which might be thought of as physical or genetic attributes that cannot be modified by an athlete, similar to height or wingspan. Also, just as in sports, practice improves the skills of the players similar to traditional athletes.

One of the differences between esports and traditional sports lie in the collection of the data. In theory esport data is “perfect,” meaning that some concerns with sports data don’t exist in esports. Also, athletes are younger which could impact player projections. Furthermore, unlike established sports, conditions of the game can change at any time. For example, new patches can be added that potentially drastically change the roles, making season to season comparisons difficult.

Overall, though, esports analytics, as we largely show, are similar enough to sports analytics that it makes sense to treat them similarly.

3 Data Overview

Using the RDota2 library in R (Boutaris, 2016), we downloaded 4,758 rows of data describing professional level Dota 2 games from Valve's Steam API (`https://steamcommunity.com/dev`). The games were played between January and October 2018 and contained 885 unique players with 10 or more games played. For each game, the API allowed us to capture a variety of player-specific data.

The following is a description of API-supplied variables, along with the abbreviations that we use for some of the statistics:

- game = Game ID: a unique identifier for professional game
- player = Player ID: a unique identifier for player
- id = Hero ID: character with which the user played
- k = Kills: number of Kills per player (higher is better)
- d = Deaths: number of deaths per player (lower is better)
- a = Assists: number of assists per player
- lh = LastHits: number of times a player lands the final hit on an enemy or object. This is a move that adds gold and removes enemies or objects from the battlefield. (higher is better)
- dn = Denies: number of denies. A deny is when you keep your opponent from landing a last hit by sacrificing a creep under your control.(higher is better)
- gpm = Gold: gold farmed per minute (higher is better)
- xpm = Exp: experience gained per minute (higher is better)
- lev = Lev: level = Hero level at the end of the game (higher is better)
- w = Win: TRUE if team won game, FALSE if lost

Throughout this manuscript we assume that the data is free from measurement error or recorder bias. One would assume that esports data would avoid these biases as the system should be able to record exactly what happened without error. As mentioned above, compared to the inherently imperfect collection mechanisms used in traditional sports, we are confident that the data used in this study is at least as reliable as data from traditional sports.

4 Gold as a continuous proxy for wins

Unlike traditional sports, Dota 2 lacks a true scoring mechanism. Without a score, it would be tempting to consider wins and losses as binary outcomes for modeling purposes. While it is certainly possible to model wins as a binary variable, these types of models tend to lose information when compared to scoring-based models that attempt to model each team's score, selecting the team with the highest score as the winner, (Ganguly and Frank, 2018). For example, capturing "blowout" wins as opposed to close games is impossible without a scoring metric.

At the player level, having a scoring mechanism is critical to assessing individuals' contributions to the team. This paper will focus on a version of the plus/minus statistic, which is impossible to determine without a scoring metric at the player level.

In Dota 2, gold is defined as, "the currency used to buy items or instantly revive your hero" and it is, "earned from killing heroes, creeps, or buildings" (dota2.gamepedia.com, 2019). By design, gold is awarded to players for doing things that help their team win, so it is a natural proxy for scoring. In this way, gold could be considered similar to metrics in traditional sports that are highly correlated with winning (e.g., shots in hockey or offensive production in football), but gold is also unique in that it is used to purchase items in the game that further help teams.

Unlike traditional sports metrics that tend to increase as teams play well, gold is actively pursued because it helps teams win. Because gold is not a perfect analogy to scoring, we will use the remainder of this section to establish the relationship between earning gold and performance.

4.1 Team-level relationship between Gold and Wins

Perhaps the simplest test for the applicability of gold as a scoring mechanism is examining the relationship between gold differential and wins. If gold is a good scoring mechanism, the winning team should nearly always earn more gold than the losing team. Because winning is only relevant at the team-level (players win or lose as a team), we examine the relationship between team total gold earned and wins.

The relationship between gold (specifically, gold differential) and wins is demonstrated in Figure 1. When compared to the other statistics, gold seems to be an especially good discriminator in that achieving a higher gold differential is strongly associated with winning.

Figure 2 further demonstrates the strength of this relationship, especially in comparison to the other available statistics. The first pane shows that in 97% of games, the team that scored more gold won. The second pane shows the Area Under Curve (AUC) metric which should be interpreted as: the probability of correctly classifying a randomly chosen win and a randomly chosen loss is 99.5%. The third pane shows LogLoss, another measure of classification accuracy that penalizes wrong classifications that are "confident" (i.e., they classify with high probability). Lower LogLoss is associated with better discrimination.

All three of these metrics show that gold is a strong discriminator of wins in general, and it is certainly the best discriminator of the available statistics.

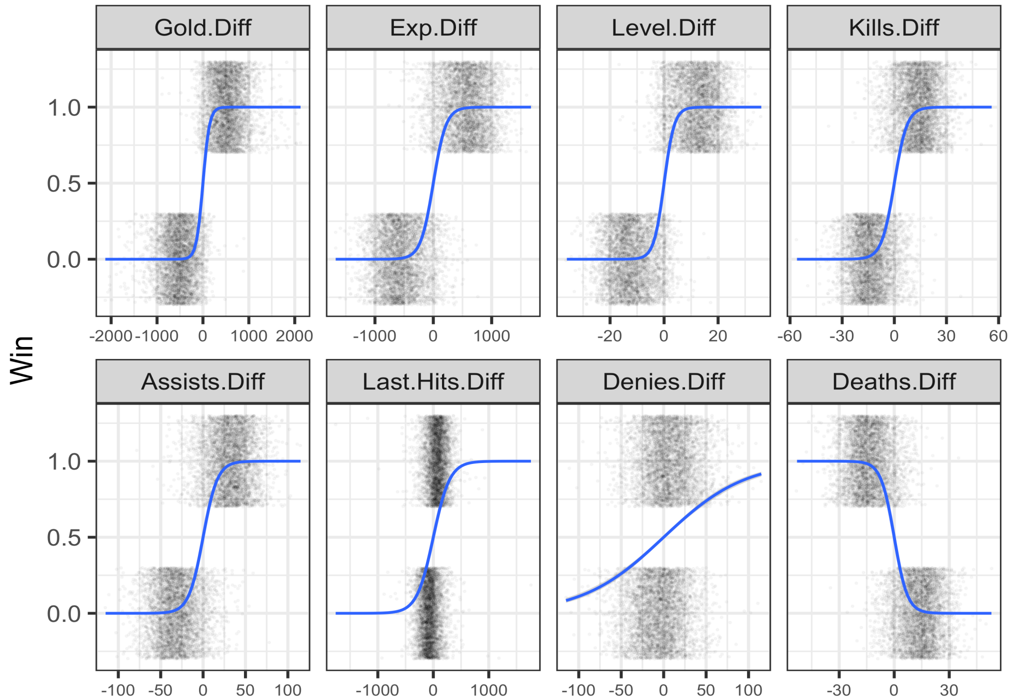

Figure 1: Scatter plot of wins vs team statistics, along with best fit logit curves. Gold appears to be the most strongly related to wins.

4.1.1 Stability of predictive ability

The above section describes how gold seems to be a good proxy for team success; however, it is important to consider the stability of these metrics before applying them. In the first sections of this paper, we describe how the data contains many players with only a few games played, so it will be especially important to determine a minimum number of games necessary to assume some stability in team statistics. We would also like to explore how team box score statistics in past games are related to team box score statistics in future games. In particular, we would like to estimate the stability and predictive ability of each statistic by exploring correlations among the statistics for different numbers of games. We can get a feel for how many games it takes for a stat to become stable, or how many games it takes before it's a good predictor. We are interested in answering the questions:

- **Stability.** What is the correlation between team stats in n games and team stats in another n games
- **Predictive ability.** What is the correlation between team stats in n games and Gold in another n games. Also, how closely associated are team stats in n games and Win% in another n games?

In Figure 3 we show the stability (left) and predictive ability (right) of different team statistics for values of n between 1 and 50. It may seem counter intuitive at first that the Denies and Last Hits are the most stable, until we recall that those were the statistics that were the least associated with winning. If we are comparing one 10-game sample to another, and the samples we choose have a different number of wins, we would expect that to have an impact on the statistics that are associated with winning, and possibly lower those correlations, but have no impact on the statistics that are not associated with winning. So it's not surprising then that Denies and Last Hits are the most stable. Note also that Gold is the most stable among the statistics that are strongly associated with winning. While Gold appears important as a team statistic, it may also be of interest to determine whether it is associated with individual player performance and other individual player statistics.

As an alternative to predicting Gold in out-of-sample games, we can also show how well these team statistics predict Win% in out-of-sample games. We built a single variable logistic regression model for each team statistic. In each model, the predictor is the team statistic in n games, and the outcome is Win% in another n games. For comparison, we also built a model using Win% in n games as a predictor, with Win% in another n games as the outcome. In Figure 4, we show the Log Loss of these models. For each value of n, the Log Loss has been centered and scaled so that all lines and values of n can fit easily on the same figure. Note that Gold Differential (thick black line) has the lowest Log Loss, indicating that the best predictor of Win% in out of sample games is Gold Differential. In particular, Gold Differential is better than Win% (thick maroon line) at predicting out-of-sample Win%. This gives alternative evidence for using Gold to evaluate and predict the performance of teams.

4.2 Relationships among player statistics within games

In Figure 5, we show the within-game correlation matrix for player statistics. In general, earning gold is associated with positive contributions such as experience, kills, and denies while gold and deaths appear to be negatively correlated. This is evidence that gold is a good measure of in-game performance and is ultimately a reasonable scoring metric.

Notably, assists do not seem to be strongly correlated with gold even though they represent a positive contribution. This is likely due to the fact that assists are not necessary for kills to occur, so their presence merely suggests that it took more than one player to complete a kill. In other words, a player who scores many assists may have a different play style that is not strongly beneficial or detrimental when compared to other play styles. This idea is further supported by the fact that assists are positively correlated with deaths, indicating that some players may fall into more supportive roles wherein they score few kills themselves, but weaken the opposing players.

In summary, we have given both qualitative and quantitative reasons for using gold as a continuous variable that is a proxy for performance of a team or player. Gold is the statistic that is the most correlated with other positive statistics, gold is the statistic that stabilizes the most quickly, and gold is the statistic that is the best discriminator of wins. But perhaps most importantly, the game was designed so that when players do good things, they get gold, and it is in a sense a catch-all statistic for positive actions by a player. For these reasons, will use gold as a continuous proxy for wins throughout the remainder of this paper.

5 Regression-based player metrics

We next create a series of regression models for evaluating individual player contributions. First, we model player contributions directly to wins. Then, we demonstrate how an Adjusted Plus Minus (APM) approach using gold as a proxy for scoring provides significant performance improvements.

5.1 Bayesian logistic regression model using Wins as the outcome

Using the previously described data from 4578 professional Dota 2 games we use information about the players as well as the team (Radiant or Dire) played, as well as the outcome of the game. If a player played in fewer than 10 games we call the player a "replacement" level player, resulting in 885 unique players in the game. Let $j \in \{1, 2, \cdots, 886\}$ index the players, we define, for $j \neq 886$:

$$X_j = \begin{cases} 1, & \text{player } j \text{ is playing Radiant in game} \\ -1, & \text{player } j \text{ is playing Dire in game} \\ 0 & \text{player } j \text{ not playing in game} \end{cases}$$

For $j = 886$ we define $X_{886} = k$ where k is the number of replacement players playing Radiant minus the number of replacement players playing Dire.

As an example, if players 1-5 play for Radiant, players 6-8 play for Dire and Dire has two replacement players, the row of X would look like

$$X = [1\ 1\ 1\ 1\ 1\ -1\ -1\ -1\ 0\ 0 \cdots 0\ -2]$$

Letting y be 1 if Radiant was victorious and 0 if Dire was victorious, our model to predict win probabilities can be written as

$$\text{logit}\ \pi = \mu + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{886} X_{886} \tag{1}$$

$$y \sim \text{Bernoulli}\ (\pi) \tag{2}$$

The intercept μ can be interpreted as an estimate of the Radiant advantage in the game, where a positive value of μ indicates that the odds ratio between a team playing as Radiant or Dire can be estimated by $\exp(\hat{\mu})$. This is analogous to a home team advantage.

In order to fit this model, we use a Bayesian methodology that requires stipulating prior distributions on all parameters in the model. To this end, letting N^+ be the half-normal density we let:

$$\mu \sim N(0,2) \tag{3}$$

$$\beta_j \sim N(0,\sigma_b) \tag{4}$$

$$\sigma_b \sim N^+(0,1) \tag{5}$$

The choices of these priors is justified as follows: while there is a known advantage for choosing Radiant over Dire, see e.g. (KC, 2017), it seems unrealistic that this advantage results in a greater than 10 times advantage for the odds, or equivalently μ is most likely is less than 2.3. Therefore, a completely uninformative prior for μ seems unrealistic in the context of the problem. We are less constrictive in our prior selection for β_j as we assume little prior knowledge on the skill of the players. Therefore we let $\beta_j \sim N(0, \sigma_b)$. We treat σ_b as a parameter to be estimated by the data. Essentially, we have constructed a hierarchical model where each player-varying parameter is drawn from a common distribution. If there is no difference between players we would expect $\sigma_b \to 0$. Whereas by not placing an intentionally vague prior on σ_b such as an improper prior on the real-line, we allow for some pooling of information between players to reflect the fact that there are similarities between all players who have made it to the professional level of Dota 2. By placing a half-normal with scale parameter 1 on σ_b we are making the assumption that the players are relatively homogeneous. Therefore, differences that we observe between our β values can likely be attributed to true differences between the individuals.

The model given by (4) is similar to the model for player performance in basketball given in (Deshpande and Jensen, 2016). As in basketball, there is a high degree of collinearity requiring a regularization prior. The choice of a $N(0, \sigma_b)$ prior corresponds to an L-2 penalty, whereas (Deshpande and Jensen, 2016) chose a Laplace prior which shrinks some of the β coefficients to exactly zero. We adopted the former in order to obtain coefficient values for each player similar to work done in hockey using ridge regression (Macdonald, 2011).

Results of the wins model

In order to conduct inference for the model, we use Stan, (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddell, 2017), through the R interface, Rstan, (Stan Development Team, 2018). To fit the model, we used three Markov Chain Monte Carlo (MCMC) chains drawn from the posterior distribution. We used a burn in of 2500 iterations per chain, then sampled 2500 iterations from the three chains. In order to determine if the chains had converged we monitored the $\hat{R}$ values which were all ≈ 1 and observed that the number of effective samples were all sufficiently high. Further, there were no divergent transitions nor other indicators of improper sampling from the posterior. Stan code and code to fit the model can be found on our Github repository at (GitHub link removed to preserve anonymity during the review process, but will be added before final publication).

Next, we conducted inference for (2). Following convergence we obtained 7500 samples from the posterior distribution for all β values as well as σ_b and μ. A subset of the results are given in Table 1.

Table 1: Coefficients for top three players from model (2) with radiant advantage

Meaning	Parameter	Posterior Mean	Posterior Standard Deviation
Radiant Advantage	μ	0.175	0.034
Kharis "SkyLark" Zafeiriou	β_{526}	0.55	0.31
Anathan Pham	β_{552}	0.54	0.32
Anucha "Jabz" Jirawong	β_{6}	0.53	0.35
Vladimir "RodjER" Nikogosyan	β_{346}	0.53	0.39
Replacement Player	β_{886}	-0.52	0.036
Standard Error for β	σ	0.39	0.001

According to this model, we see that the odds of a team winning if Kharis "SkyLark" Zafeiriou is on their team increase by a factor of 2.91 over having a replacement player on the team. The top three players identified by the model are Zafeiriou, Anathan Pham, and Vladimir Nikogosyan, whose odds of winning increase above a replacement player by 2.91, 2.88, and 2.85 respectively.

5.2 Bayesian linear regressions model using gold as the outcome

One downside of the model given in (2) is that it does not consider a metric for offensive vs. defensive contributions separately, rather it only gives the probability for a team winning depending who is playing. Furthermore, this model does not adjust for the quality of the opposition. It is desirable to have a metric that determines the impact of a player while controlling for the performance of his or her teammates and opponents. In traditional sports statistics, this is often done using adjust plus-minus metrics based on, for instance, goals in hockey (Macdonald, 2011) or points in the NBA (Kubatko, Oliver, Pelton, and Rosenbaum (2007) or Fearnhead and Taylor (2011)). While Dota 2 lacks a true scoring mechanism, we demonstrated in Section 2 that gold is a useful proxy. As demonstrated below, using gold as an outcome allows for player metrics that are better predictors of future performance.

Prior to stating our model we make the following definitions:

$$i = \text{game} \tag{6}$$

$$t = \text{team 1 or 2} \tag{7}$$

$$g_{i,t} = \text{centered gold obtained in game } i \text{ by team } t \tag{8}$$

$$j = \text{player } 1, \cdots, 885 \tag{9}$$

$$\gamma_{i,j} = 1 \text{ if player } j \text{ is on offense, 0 otherwise} \tag{10}$$

$$\zeta_{i,j} = 1 \text{ if player } j \text{ is on defense, 0 otherwise} \tag{11}$$

$$r_{i,t} = 1 \text{ if offensive team } t \text{ is playing Radiant in game } i\text{, else 0} \tag{12}$$

To improve fit, we center the $g_{i,t}$ term by taking the raw gold obtained by team in a game and subtracting off the overall average gold professional teams score per game.

Note that Dota 2 is similar to hockey as the teams rapidly shift between being on the offense and being on the defense. To model this, each game played has two entries in the data set, one for the gold that they score and one for the gold that is scored against them. To clarify this, say that one of the teams, arbitrarily referred to as team 1, consists of players 1,2,3,4,5 playing Radiant and the other team (team 2) consists of players 6,7,8,9,10 playing Dire. Here $g_{i,1}$ is equal to the gold that team 1 obtained in game i, and in this case $\gamma_{i,1}$ through $\gamma_{i,5}$ are equal to 1, $\zeta_{i,6}$ through $\zeta_{i,10}$ are equal to 1, and $r_{i,1} = 1$. The second entry is when team 1 is on defense. Now $g_{i,2}$ is equal to the gold that team 2 obtained in game i. For this entry $\gamma_{i,6}$ through $\gamma_{i,10}$ are 1 and $\zeta_{i,1}$ through $\zeta i, 5$ are equal to 1 and $r_{i,2} = 0$ as the offensive team is now Dire. We show two rows for the predictor matrix X and the outcome y for this example game in (13).

$$X = \left[r_i \; \middle| \; \gamma_i \; \middle| \; \zeta_i \right] = \left[\begin{array}{c|ccc|ccc} 1 & 1\,1\,1\,1\,1 & 0\,0\,0\,0\,0 & 0 \cdots 0 & 0\,0\,0\,0\,0 & 1\,1\,1\,1\,1 & 0 \cdots 0 \\ 0 & 0\,0\,0\,0\,0 & 1\,1\,1\,1\,1 & 0 \cdots 0 & 1\,1\,1\,1\,1 & 0\,0\,0\,0\,0 & 0 \cdots 0 \end{array} \right]$$

$$\text{y} = \begin{bmatrix} g_{i,1} \\ g_{i,2} \end{bmatrix} \begin{array}{c} \text{(Radiant Gold)} \\ \text{(Dire Gold)} \end{array} \tag{13}$$

As in the logistic regression, we use a Bayesian inferential model and make the following distributional assumptions:

$$\mu_{i,t} = \gamma_i \beta + \zeta_i \kappa + r_{i,t} \rho \tag{14}$$

$$g_{i,t} \sim N(\mu_{i,t}, \sigma_g) \tag{15}$$

In (15) note that γ_i and ζ_i are row vectors with non-zero entries for who is playing in the game and ρ is a scaler that captures the known radiant advantage similar as to what is done in (2). β and κ are column vectors of parameters that define the offensive and defensive strength of the players, respectively.

Finally, we set prior distributions as:

$$\beta_j \sim N(0, \sigma_b) \tag{16}$$

$$\kappa_j \sim N(0, \sigma_k) \tag{17}$$

$$\sigma_g \sim t^+(5, 0, 10) \tag{18}$$

$$\sigma_b \sim t^+(5, 0, 10) \tag{19}$$

$$\sigma_k \sim t^+(5, 0, 10) \tag{20}$$

$$\rho \sim N(0, 100) \tag{21}$$

$$\tag{22}$$

where in (18), $t^+(d, c, s)$ is a half-t distribution with d degrees of freedom, centered at c with scale parameter s. As opposed to model (4), we have limited prior intuition about the range of our variance parameters. The half-t prior has heavier tails than the half-normal and allows for larger deviations from 0 for the posterior of σ_b, σ_k and σ_g than the model given in (4). We place the same prior independently on β_j and κ_j, that is, we assume the impacts of defense and offense on Gold are relatively comparable.

The above model can be conceptualized as a generalization of a ridge regression model that has been previously shown to improve upon the traditional plus minus metric in ice hockey (Macdonald, 2012). Both ridge regression and the model given in (15) are models that are sometimes referred to as regression shrinkage models (see e.g. Tibshirani (1996)). The ridge regression model, however, fixes the variance of the prior distribution for β and κ using internal cross validation. The model presented in (15) accounts for the uncertainty in this by putting a further hyperprior distribution on the variance of β and κ. In both cases, though, the collinearity that is often present in this data is mitigated (e.g. Walker and Birch

(1988)).

5.2.1 Results from the Gold model

We again use Stan to conduct inference using three MCMC chains of 5000 iterations letting the first 2500 serve as a burn-in. In order to determine convergence we again monitor $\hat{R}$, number of effective samples, as well as ensuring there were no divergent transitions within the chains. In our sampling we also calculate a measure of gold gained above replacement by monitoring $\beta_j - \beta_{886}$ and gold prevented above replacement by monitoring $\kappa_{886} - \kappa_j$. Code for fitting the model and sampling from the posterior can be obtained from (GitHub link removed to preserve anonymity during the review process, but will be added before final publication).

This model allows us to define offensive players as those whose presence in the game contributes to gold gained for the team whereas a defensive player is one whose presence prevents the opposing team from obtaining gold. Using the model parameters from above the top offensive players, then, can be considered the players who have the highest mean $\beta_j - \beta_{886}$ value and the top defensive players can be found through examining the players who have the highest $\kappa_{886} - \kappa_j$ value. The top 5 offensive players are given in Table 2 and the top 5 defensive players are given in Table 3.

Table 2: Coefficients for top offensive players in Gold.

Top Offensive Players	Parameter	Gold Above Replacement (GAR)	Standard Deviation of (GAR)
Kharis “SkyLark” Zafeiriou	$\beta_{526} - \beta_{886}$	178.4	45.4
Zhou “James” Yifu	$\beta_{351} - \beta_{886}$	162.8	46.5
ClamsCasino	$\beta_{796} - \beta_{886}$	162.3	51.6
Anucha “Jabz” Jirawong	$\beta_{6} - \beta_{886}$	162.2	48.1
“Maybe” Team LGD	$\beta_{65} - \beta_{886}$	162.0	49.0

Table 3: Coefficients for top defensive players in Gold.

Meaning	Parameter	Posterior Mean	Posterior Standard Deviation
“The Jokes” Team Hippomaniacs	$\kappa_{886} - \kappa_{708}$	-182.6	55.5
ClamsCasino	$\kappa_{886} - \kappa_{603}$	-167.1	52.1
Vladmir “NoOne” Minenko	$\kappa_{886} - \kappa_{19}$	-164.6	49.3
Vladmir “RodjER” Nikogosian	$\kappa_{886} - \kappa_{346}$	-158.9	50.0
Guvara - DeathBringer Gaming	$\kappa_{886} - \kappa_{515}$	-153.9	47

In Figure 6, we give a scatter plot of the posterior mean offensive and defensive coefficients for each player. As we observe, on average professional players outperform replacement player either offensively or defensively. Note that in this figure the axis represents the expected offensive and defensive contribution of a replacement player.

In Figure 7, we show examples of distributions for five players,

- the best overall player (black)
- the best defensive player (green)
- the best offensive player (dark pink)
- the worst overall player (purple). Also the worst offensive player.
- the worst defensive player (light pink)

along with the mean (large point) and 95% credible region (ellipse). Note that the credible regions for, for example, the best and worse players do not overlap, however, there is large overlap in the credible regions for many players, including players 89128606 (“Clams Casino”), 82709828 (Mirek “Smurf” Bican), and 27178898 (“Skylark”). We note that in most cases, the uncertainly in the offensive coefficient is smaller than the uncertainty in the defensive coefficients, though since they are fairly close (average standard deviations 48.5 and 51.7 respectively), the credible regions appear to be roughly circular as opposed to elliptical.

While the majority of players identified as top offensive and top defensive players would be familiar to fans of Dota 2, such as “Maybe” from team LGD or “No One” from Virtus Pro, there is one name that jumps out as a potentially oddity, that is the player going by the moniker ClamsCasino. Subsequent analysis determined that this player only played in 12 games, making them almost in the class of replacement players, and the games he or she played in resulted in blowouts for the victors. While certainly possible that a larger set of professional games might knock this individual out of the top 5, it is also possible that this is an up and coming player who managers of larger professional teams may want to further scout. Additionally, without performing identity resolution, we accept it is possible that professional players may use alternate named accounts for various reasons.

5.2.2 Box stats vs APM coefficients

In Figure 8, we show correlations between the results our models (along the side) and player box score statistics (along the top). We note that the offensive (off) and defensive (def) coefficients from the gold model are correlated with gold for and gold against, which is expected since those can be considered unadjusted versions of the coefficients.

We also note that assists are positively correlated with the offensive coefficient, which we would expect if assists are positive contributions to the team. In Section 2 we discussed the somewhat counter-intuitive result that players’ assists were negatively correlated with their personal gold, and suggested that the reason was likely due to the different roles that players fill on their teams. Here we have the more intuitive result that assists are positively correlated with the impact that the player has on the team’s gold. This suggests that a player’s model-based metrics are a more comprehensive representation of how a player impacts the game than the player’s gold statistic alone.

6 Predicting game outcomes

In order to determine how this model performs as a predictive model, we downloaded 300 professional Dota 2 games from November of 2018, the month immediately following the training data used in creating the above model. Note that these games are considered out of sample games as they were not used to inform either of the models presented above. For each game, we included the game in our predictive pool if neither team was entirely composed of "replacement players." That is, we included the game if both teams had members corresponding to one of our β values. Of the 300 professional games we looked at, 150 of the games met the criteria described above. For each game in our sample, we used our found posterior distributions for the β values associated with the players as well as the posterior distribution for μ and found the probability of Radiant winning that game. For example, for the first game in the sample, Theeban Siva, MoonMeander, Xepher, InYourDream, Ahjit played as Radiant. This likely is the professional Dota team known as the Tigers. SoNNeikO, Crystallize, Blizzy, MagicaL, and I "heart" ILTW, played as Dire. This likely is the professional Dota team Natus Vincere. The players playing correspond to β_{184}, β_{215}, β_{401}, β_{607}, β_{769} and have an x_j of 1, and β_{126}, β_{150}, β_{151}, β_{373}, β_{490} and have an x_j of -1. We then take samples from the posterior distribution of each β value and μ and derive the predictive density of the probability of Radiant winning. For example, as we see in Figure 9, there is significant uncertainty in the outcome between Natus Vincere and the Tigers, however the model slightly favors the Tigers playing as Radiant.

The median of the density is 0.55 suggesting we would predict Radiant would win, but with low confidence. In this game, Radiant was victorious suggesting that if we classified to the median prediction we would have been correct.

For all 150 games where we had knowledge of players on both teams, we found the probability of both Dire and Radiant winning the match and grouped them by the posterior medians. For example, we see in Table 4 that there were 21 games where the model gave either the team playing Dire or the team playing Radiant a 70-75% probability of winning the game. In the actual outcome of these games, 16 of those games were won by the favored team, or 76% of the time the team favored won. The results here are clearly encouraging in a straightforward model being able to give pre-game probabilities of teams winning the game.

Prob of Radiant or Dire Winning	Number of Games	Number of Times Won	Actual Percent
50-55%	21	12	57%
55-60%	31	20	64%
60-65%	29	11	38%
65-70%	11	7	64%
70-75%	21	16	76%
75-80%	12	8	67%
80-85%	9	7	78%
85-90%	9	8	89%
90-100%	7	7	100%

Table 4: How Model Performs on Future Games

The same out of sample data was also fit to model (15) giving expected gold for both Dire and Radiant. If the expected gold for for Dire was higher than the expected gold for Radiant, we subsequently predicted that Dire would win. As seen in Table 5, as the predicted differences in gold increases, the percentage of games that the model correctly predicts also increases. This further gives support to the idea that gold can serve as a scoring metric and a higher expected gold differential can serve as a measure of team strength in analyzing the difference between two teams.

Currently, the most common way to rank players is either through total dollars won in competitions or through box score statistics, such as highest average gold per game. For example, the website www.opendota.com offers rankings by duration, kills, deaths, assists, gold, XPM, and other scores. As explained above,

Expected difference in Gold	Number of Games	Percent Correctly Predicted
Δ Gold >0	150	60%
Δ Gold >100	92	65%
Δ Gold >200	60	70%
Δ Gold >300	42	76%
Δ Gold >400	30	80%
Δ Gold >500	15	93%
Δ Gold >600	6	100%

Table 5: How the Gold model performs on future games. As the expected difference in gold increases, the percent correctly predicted also increases

of all the box score statistics commonly used to rank players, gold has the strongest correlation with winning. Therefore, a box score method (a reasonable naive model) of determining the winning team is to sum the average gold for each side and predict the winner based on the team with the highest expected gold. Using this methodology, we found the expected gold for each team within the same 150 professional Dota 2 games from November of 2018. As seen in Table 6 while using box scores to predict the winning team results in a similar performance to the model based approach (57% vs 60%), the box score approach does a poor job in quantifying how much better a team is than another team. Of the 150 games in the sample, only 37 had an expected difference in gold over over 200, and of those 37, the box score statistics only correctly classified 54% of them. On the other hand, the Bayesian APM model gave an expected gold differential of greater than 200 in 60 of the cases, and of those 60 correctly classified 70% of them.

The Bayesian APM model is also much more stable than the box score model. In other words, if we predict several months in the future using the Bayesian APM model, it will perform similarly as the one month ahead prediction performed above. To demonstrate this stability we downloaded 2356 professional games from January to March of 2019 and used both the Bayesian APM as well as the box score methodology to predict expected difference in gold. We then sorted the predictions

Expected difference in Gold	Number of Games	Percent Correctly Predicted
Δ Gold >0	150	57%
Δ Gold >100	89	62%
Δ Gold >200	37	54%
Δ Gold >300	11	73%
Δ Gold >400	2	100%
Δ Gold >500	2	100%
Δ Gold >600	1	100%

Table 6: Box Scores Only

from both models from highest expected difference in gold between teams to lowest expected difference in gold. Starting with the highest expected difference we then calculated the cumulative proportion of correct predictions. In Figure 10, as we move left to right on the X-axis the expected difference in gold decreases.

We would, in general, expect the cumulative proportion of correct predictions to start off high, as a higher expected difference in gold should correspond to a higher level of certainty that one team is better than the other, and decrease as the expected difference in gold decreases. In general, if expected difference in gold is close to zero the models offer little to differentiate the two teams. We would also expect the plots to have higher variability and level out as the number of games predicted gets bigger. As seen in Figure 10, we see that the Bayesian APM using gold (black) as a response consistently outperforms the model using box scores only (pink).

We have included results when using the coefficients from the Win model in this figure as well (teal). The player coefficients from the Gold model perform comparably to the coefficients from the Wins model in terms of predicting out-of-sample wins for future games involving those players, despite the fact that the Gold model was built to predict Gold as opposed to Wins. The Gold model also has the added benefit over the Wins model because it results in separate offensive and defensive coefficients. Both regression models perform much better than box score version of Gold at predicting future wins, and this provides strong evidence that advanced statistics like these can provide better measures of a player's true ability in esports, in the same way they help with player evaluation in traditional sports.

7 Conclusion

In this manuscript we demonstrated how traditional sports analytics can play a role in the analysis of esports. In particular, we showed how gold in Dota 2 can serve as a proxy for goals or points. Using gold as a scoring metric, we quantified players' offensive and defensive contributions and constructed a Bayesian APM model. We used this model to further refine the offensive and defensive player metrics by adjusting for the strength of opponents and teammates. This analysis offers insights useful for team managers and sponsors to identify both top performers as well as upcoming stars.

The Bayesian APM model also outperforms our naive, box score-based statistics as a predictive model. The 66% accuracy obtained on the set of 2356 games from January to March of 2019 by using the APM model rivals top performers in traditional sports as seen in (Spann and Skiera, 2009). However, it should be noted that more work should be done to assess the model's predictive ability as it only accurately predicted 60% accurately on a smaller set of 150 out of sample games. Furthermore, there are likely interactions that occur between players that are not accounted for in the linear model that might be better accounted for using machine learning techniques. We also did not adjust for hero data. As explained in the introduction, there are, at the time of writing, well over 100 heroes that an individual can choose from. The choice of hero likely impacts both the tasks the individual is performing as well as their performance. There is an abundance of interesting analyses that one could perform using hero data. For example, in (Ivan

Ramler and Choong-Soo Lee, 2018), the authors investigate synergies among characters in League of Legends. These same types of analyses could be done for Dota 2 in the development of future models.

Overall, the application of analytical techniques borrowed from traditional sports to esports is in the early stages. We anticipate that there is much more that could be done that we have not yet considered and we encourage analysts who traditionally work with conventional sports data to consider expanding to this growing field.

Acknowledgements

The authors would like to thank the reviewers and editors for several useful comments that helped improve this work.

References

Boutaris, T. (2016): *RDota2: An R Steam API Client for Valve's Dota2*, URL `https://CRAN.R-project.org/package=RDota2`, r package version 0.1.6.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017): "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76.

Deshpande, S. K. and S. T. Jensen (2016): "Estimating an NBA Players Impact on His Teams Chances of Winning," *Journal of Quantitative Analysis in Sports*, 12, 51–72.

DOTA2.com (2019): "The International Championships," `http://www.dota2.com/international/overview/`, accessed: 2019-01-09.

dota2.gamepedia.com (2019): "DOTA 2 Wiki - Gold," `https://dota2.gamepedia.com/Gold`, accessed: 2019-07-10.

Drachen, A., M. Yancey, J. Maguire, D. Chu, I. Y. Wang, T. Mahlmann, M. Schubert, and D. Klabajan (2014): "Skill-Based Differences in Spatio-Temporal Team Behaviour in Defence of the Ancients 2 (Dota 2)," in *2014 IEEE Games Media Entertainment*, IEEE, 1–8.

Eggert, C., M. Herrlich, J. Smeddinck, and R. Malaka (2015): "Classification of Player Roles in the Team-Based Multi-Player Game Dota 2," in *International Conference on Entertainment Computing*, Springer, 112–125.

Fearnhead, P. and B. M. Taylor (2011): "On Estimating the Ability of NBA Players," *Journal of Quantitative Analysis in Sports*, 7.

Fisher, E. (2017): "Colleges Offer Analytics Track," `https://www.sportsbusinessdaily.com/Journal/Issues/2017/06/05/In-Depth/Data-side.aspx`, accessed: 2019-09-11.

Ganguly, S. and N. Frank (2018): "The Problem with Win Probability," in *Proc. of the 12th MIT Sloan Sports Analytics Conf.*

Gramacy, R. B., S. T. Jensen, and M. Taddy (2013): "Estimating Player Contribution in Hockey with Regularized Logistic Regression," *Journal of Quantitative Analysis in Sports*, 8, 97–111, URL `http://www.degruyter.com/view/j/jqas.2013.9.issue-1/jqas-2012-0001/jqas-2012-0001.xml`.

Ilardi, S. (2007): "Adjusted Plus-Minus: An Idea Whose Time Has Come," `http://www.82games.com/ilardi1.htm`.

Ilardi, S. and A. Barzilai (2008): "Adjusted Plus-Minus Ratings: New and Improved for 2007-2008," `http://www.82games.com/ilardi2.htm`.

Ivan Ramler and Choong-Soo Lee (2018): "Identifying Symbiotic Relationships between Champions in League of Legends," In progress.

Kaytoue, M., A. Silva, L. Cerf, W. Meira Jr, and C. Raïssi (2012): "Watch Me Playing, I am a Professional: A First Study on Video Game Live Streaming," in *Proceedings of the 21st International Conference on World Wide Web*, ACM, 1181–1188.

KC (2017): "Dota 2: Here's Why Radiant Has A Lot Of Advantage Than The Dire," http://www.itechpost.com/articles/77468/20170125/dota-2-heres-why-radiant-has-a-lot-of-advantage-than-the-dire.htm, accessed: 2018-12-13.

Kubatko, J., D. Oliver, K. Pelton, and D. T. Rosenbaum (2007): "A Starting Point for Analyzing Basketball Statistics," *Journal of Quantitative Analysis in Sports*, 3.

Macdonald, B. (2011): "A Regression-Based Adjusted Plus-Minus Statistic for NHL Players," *Journal of Quantitative Analysis in Sports*, 7.

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.

Marshall, A. (2019): "2018 US Open Prize Money to Reach $53 Million," https://www.usopen.org/en_US/news/articles/2018-07-17/2018-07-16_us_open_prize_money_to_reach_record_53_million.html.

Matano, F., L. F. Richardson, T. Pospisil, C. Eubanks, and J. Qin (2018): "Augmenting Adjusted Plus-Minus in Soccer with FIFA Ratings," *arXiv preprint arXiv:1810.08032*.

Okamoto, D. M. (2011): "Stratified Odds Ratios for Evaluating NBA Players Based on their Plus/Minus Statistics," *Journal of Quantitative Analysis in Sports*, 7.

PGA.com (2019): "2018 PGA Championship: Purse and Winner's Share," https://www.pga.com/events/pgachampionship/2018-pga-championship-purse-and-winners-share, accessed: 2019-01-09.

Rosenbaum, D. (2004): "Measuring How NBA Players Help Their Teams Win," http://www.82games.com/comm30.htm.

Schuckers, M. and J. Curro (2013): "Total Hockey Rating (THoR): A Comprehensive Statistical Rating of National Hockey League Forwards and Defensemen Based Upon All On-Ice Events," URL http://statsportsconsulting.com/main/wp-content/uploads/Schuckers_Curro_MIT_Sloan_THoR.pdf.

Semenov, A., P. Romov, S. Korolev, D. Yashkov, and K. Neklyudov (2016): "Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota 2," in *International Conference on Analysis of Images, Social Networks and Texts*, Springer, 26–37.

Sill, J. (2010): "Improved NBA Adjusted+/-Using Regularization and Out-of-Sample Testing," in *Proceedings of the 2010 MIT Sloan Sports Analytics Conference*.

Spann, M. and B. Skiera (2009): "Sports Forecasting: a Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters," *Journal of Forecasting*, 28, 55–72.

Stan Development Team (2018): "RStan: the R interface to Stan," URL http://mc-stan.org/, r package version 2.17.3.

steamcommunity.com (2019): "STEAM Community Webpage," https://steamcommunity.com/app/570, accessed: 2019-01-07.

Thomas, A. C., S. L. Ventura, S. T. Jensen, and S. Ma (2013): "Competing Process Hazard Function Models for Player Ratings in Ice Hockey," *The Annals of Applied Statistics*, 7, 1497–1524, URL https://doi.org/10.1214/13-AOAS646.

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Walker, E. and J. B. Birch (1988): "Influence Measures in Ridge Regression," *Technometrics*, 30, 221–227.

Wickham, H. (2017): *Tidyverse: Easily Install and Load the 'Tidyverse'*, r package version 1.2.1.

A R Code to Reproduce Results

The following code allows readers to download Dota 2 data in order to reproduce or extend results from this manuscript. Obtaining the data in the correct format requires multiple steps. The first step is to obtain the game ID from `https://www.opendota.com`. The website offers several search options, we collected the data by searching for all professional games in the given time periods and retained the game ID. An example of this can be found at our github page at (GitHub link removed to preserve anonymity during the review process, but will be added before final publication) as GameIDs.csv.

A.1 Obtaining the Data

The game IDs were then used with the R library RDota2 (Boutaris, 2016) and tidyverse (Wickham, 2017) in the following code. Note that this code requires an API key that can be obtained via `https://steamcommunity.com/dev`.

```r
library(RDota2)
library(tidyverse)
article_key <- "Your Key"
key_actions(action='register_key',value=article_key)
prof.dota.dat=read.csv("GameIDs.csv")

prof.dota.tib=prof.dota.dat
tot.games=length(unique(prof.dota.tib$match_id))
unique.games=unique(prof.dota.tib$match_id)
count=1
#Blank Data frame to hold player data per game
place.hold=data.frame(game=NA, player=NA, slot=NA, id=NA, i0=NA,
```

```
                                  i1=NA, i2=NA, i3=NA, i4=NA, i5=NA, bp0=NA,
                                  bp1=NA, bp2=NA, itemn=NA, kills=NA, deaths=NA,
                                  assists=NA, l_st=NA, la_hits=NA,
                             denies=NA, gpm=NA, xpm=NA, level=NA, win=NA)
prof.match.details=list()
for(k in 1:tot.games){
  game=get_match_details(unique.games[k])
  for(j in 1:10){
    place.hold[count,1]=unique.games[k]
    place.hold[count,(2:(ncol(place.hold)-1))]=
                                  data.frame(game$content[[1]][[j]][1:22])
    if(j<=5){
      place.hold[count,]$win=(prof.dota.tib%>%
                                   filter(match_id==unique.games[k])%>%
                                   select(win))[1,]
    }else{
      place.hold[count,]$win=(prof.dota.tib%>%
                                   filter(match_id==unique.games[k])%>%
                                   select(win))[6,]
    }
    count=count+1
  }
}

dota.working.data=as.tibble(place.hold)

dota.working.data$win=unlist(place.hold$win)
```

```
write.csv(dota.working.data,"FullData.csv")
```

```
key_actions(action='delete_key')
```

A.2 Preparing Data for Exploratory Analysis of Section 4

What follows is code to obtain data in format to conduct player statistics used in Section 4.2. This code can be found on our github repo and is called CreatePlayerData.R.

```
d = read.csv('FullData.csv')
h = read.csv('heroes.csv')
d$X=NULL
d$win=d$win-1 #Got coded 2,1 vice 1,0
d = d %>%
  rename(k=kills, d=deaths, a=assists, st=l_st,
         lh=la_hits, dn=denies, w=win, lev=level) %>%
  mutate(w=as.numeric(w), l=1-w, gp=1)
d$game.num = as.numeric(factor(d$game))
## create radiant/dire column
d$rd = NA
rows.r = d$slot %in% 0:4
rows.d = d$slot %in% 128:132
d$rd[rows.r] = 'radiant'
d$rd[rows.d] = 'dire'
```

```
head(d,2)
table(d$rd)

## rename slot.  instead of 0-4 and 128-132,
## make it home1-home5 and away1-away5
d$slot[rows.r] = paste0('radiant', d$slot[rows.r]+1)
d$slot[rows.d] = paste0('dire', as.numeric(d$slot[rows.d])-127)
## convert hero numbers to names

h = select(h, hero, id, str, agi, int, tot)
d = left_join(d,h)
head(d,2)

d = rename(d, hero.id=id)

## convert player numbers to names, if possible
d$player.id = d$player

stat.cols = c('k', 'd', 'a',  'lh', 'dn', 'gpm',
    'xpm', 'lev', 'w', 'l', 'gp', 'str', 'agi',
    'int', 'tot' )
dd = d %>% group_by(game, game.num, rd) %>%
  summarise_at(stat.cols, sum) %>%
  mutate(w=w/5, l=l/5, gp=gp/5)
head(dd)

## create for/against, differential, percentage
```

```
stat.cols.a = paste0(stat.cols, '.a')
stat.cols.d = paste0(stat.cols, '.d')
stat.cols.p = paste0(stat.cols, '.p')
dd[,stat.cols.a]=NA

### against
for(j in unique(dd$game)){
  dd[dd$game==j & dd$rd=='radiant', stat.cols.a] = dd[dd$game==j &
                                                       dd$rd=='dire'
, stat.cols]
  dd[dd$game==j & dd$rd=='dire'    , stat.cols.a] = dd[dd$game==j &
                                                       dd$rd=='radiant', stat.cols]
}

### gpm diff and perc
dd[stat.cols.d] = dd[,stat.cols] - dd[,stat.cols.a]
dd[stat.cols.p] = dd[,stat.cols]/(dd[,stat.cols]+dd[,stat.cols.a])
head(dd)

## create team ID
rad  = d %>%
    select(game, slot, player, rd) %>% filter(rd=='radiant') %>%
    spread(key=slot, value=player)
dire = d %>%
    select(game, slot, player, rd) %>% filter(rd=='dire'
) %>%
    spread(key=slot, value=player)
```

```
colnames(rad ) = c('game', 'rd', paste0('p', 1:5))
colnames(dire) = c('game', 'rd', paste0('p', 1:5))
cols = paste0('p', as.character(1:5))
for(j in 1:nrow(rad )){ rad[j,cols] = sort( rad[j,cols])}
for(j in 1:nrow(dire)){dire[j,cols] = sort(dire[j,cols])}
teams = bind_rows(rad, dire) %>% arrange(game)
unique.teams = unique(teams[,cols])
n.teams = nrow(unique(teams))

## label the unique teams
unique.teams$id = 1:nrow(unique.teams)

## add labels to team data.
unique.teams$players= paste(unique.teams$p1,
                            unique.teams$p2,
                            unique.teams$p3,
                            unique.teams$p4,
                            unique.teams$p5, sep='-')

teams$players = paste(teams$p1,
                      teams$p2,
                      teams$p3,
                      teams$p4,
                      teams$p5, sep='-')



## match based on those columns.
```

```
teams$team.id=match(teams$players, unique.teams$players)
teams$team.id=teams$team.id+10000
## add 10000 so that every team number is 5 digits.

## add team.id to the team data and player data
head(d ,2)
head(dd,2)
head(teams,2)
dd = left_join(dd, select(teams, game, rd, team.id))
d  = left_join(d , select(teams, game, rd, team.id))

## add players to team data
colnames(rad ) = c('game', 'rd', paste0('r', 1:5))
colnames(dire) = c('game', 'rd', paste0('d', 1:5))
rad.dire = full_join(select(rad, -rd),select(dire, -rd), by='game')
dd = left_join(dd, rad.dire)

## add team stats to player data
d = left_join(d, dd,
              by=c('game', 'game.num', 'rd', 'team.id'),
              suffix=c('', '.f'))

## save player and team data
filename = 'data/player.data.by.game.rds'
saveRDS(d, file=filename)
```

A.3 Creating Team Summaries of Data

The following code creates team summaries used in Section 4.1 to create team data.

This code can be found on our Github repo entitled 'CreateTeamData.R'

```
filename = 'rawdata/player.data.by.game.rds'
d = readRDS(file=filename)
head(d,2)
d$slot = gsub('home|away', '', d$slot)

## compute team totals
stat.cols = c('k', 'd', 'a', 'st', 'lh', 'dn', 'gpm','xpm', 'lev',
    'w', 'l', 'gp')
dt = d %>% group_by(game, ha) %>%
  summarise_at(stat.cols, sum) %>%
  mutate(w=w/5, l=l/5, gp=gp/5) ## w, l and gp are already team totals
  #so don't need to add 1 for each player
head(dt,2)

## get all players for each game
## get all heroes   for each game
dp = d %>%
    select(game, game.num, slot, player, ha) %>%
    spread(key=slot, value=player)
dh = d %>%
    select(game, game.num, slot, hero   , ha) %>%
    spread(key=slot, value=hero   )

## rename columns
```

```
player.cols = paste0('p', 1:5)
hero.cols   = paste0('h', 1:5)
colnames(dp) = c('game', 'game.num', 'ha', player.cols)
colnames(dh) = c('game', 'game.num', 'ha', hero.cols  )

## get unique teams
ordered.player.cols = gsub('p', 'o', player.cols)
ordered.player.cols
dp[,ordered.player.cols]=NA
head(dp)
## sort players so that we can find unique 5-player combos
## where order doesn't matter.
for(j in 1:nrow(dp)){
  dp[j,ordered.player.cols] = sort(dp[j,player.cols])
}

## find rows that aren't duplicated, and number them.
rows = which(!duplicated(dp[,ordered.player.cols]))
team.ids = dp[rows,ordered.player.cols]
team.ids$team.id=1:nrow(team.ids)

dp = merge(dp, team.ids, by=ordered.player.cols, all=T, sort=F)
dp = dp[order(dp$game.num),]

head(dp)

## rearrange columns, remove ordered.player.cols
```

```
cols = c('game', 'game.num', 'ha', 'team.id', player.cols)
dp = dp[,cols]
head(dp)
#### end of code for adding team.id ####

## merge players, team, heroes, and team stats
head(dt,2)
head(dp,2)
head(dh,2)
d = merge(dp, dh, by=c('game', 'game.num', 'ha'), all=T)
head(d)
d = merge(d, dt, by=c('game', 'ha'), all=T)
head(d,2)

## for and against.
## Copy the 'for' stats for the *home* team
##to the 'against' columns for the *away* team.
## Copy the 'for' stats for the *away* team
##to the 'against' columns for the *home* team.
## create a column that is the opposite of ha.
## Then merge on ha and ha.against.
d$ha.against = NA
d$ha.against[d$ha=='away'] = 'home'
d$ha.against[d$ha=='home'] = 'away'
head(d)
d = merge(select(d,-ha.against    ),
          select(d,-ha, -game.num),
```

```
        by.x=c('game', 'ha'        ),
        by.y=c('game', 'ha.against'), suffixes=c('.f', '.a'))
head(d)

## save team data
filename = 'data/team.data.by.game.rds'
saveRDS(d, file=filename)
```

A.4 Preparing Data for APM Models

The following code uses the ‘player.data.by.game.rds‘ file constructed above to prepare the data for use in the APM model. The code can be found on our Github repo entitled CreateAPMData.R.

```
library(tidyverse)
library(Matrix)
filename = 'data/player.data.by.game.rds'
dp = readRDS(filename)
dp$game    = factor(dp$game)

filename = 'data/team.data.by.game.rds'
dt = readRDS(filename)

## Rep players. GP for players
min.GP = 10
GP.players = dp %>% count(player) %>% arrange(desc(n))
rows = GP.players$n < min.GP
```

```
rep.players = as.character(GP.players$player[rows])
length(which(!rows))
head(rep.players)

## d with replacement players
dr=dp
rows = dr$player %in% rep.players
dr$player[rows] = 'rep'
head(dr,2)

rows1 = dp$rd=='radiant'
rows2 = dp$rd=='dire'
dp$player = factor(dp$player)
dr$player = factor(dr$player)


create.wide.for.apm = function(dp=NULL, dt=NULL, rows='game',
                                cols='player', outcomes='w',
                               off.def=T, rep=T){

  formula  = as.formula(paste0(' ~ ', rows, '+', cols))
  xh = xtabs(formula, data=dp[rows1,], sparse=T,
        drop.unused.levels = F)
  xa = xtabs(formula, data=dp[rows2,], sparse=T,
        drop.unused.levels = F)
  xh = as.data.frame.matrix(xh)
  xa = as.data.frame.matrix(xa)
```

```
games = rownames(xh)

if(off.def==F){
  x = xh - xa
  if(rep==T){x = select(x, -rep) %>%
      mutate(rep.h=xh$rep, rep.a=-xa$rep)}
  x$team = 'radiant'
}

if(off.def==T){
  col.names = c(paste0('o', colnames(xh)), paste0('d',
                  colnames(xh)), 'team')
  x = rbind(cbind(xh, xa, team='radiant'),
            cbind(xa, xh, team='dire'))
  colnames(x) = col.names
}

x$game = games

## create y

## if off.def=F, take radiant wins
if(off.def==F){
  y = dt %>%
    filter(rd=='radiant') %>%
    mutate(radiant.w=w, team='radiant') %>%
    select(game, radiant.w, team)
```

```
  }

  if(off.def==T){
    yh = dt %>%
         filter(rd=='radiant') %>%
         select('game', outcomes) %>%
         mutate(team='radiant')
    ya = dt %>%
         filter(rd=='dire') %>%
         select('game', outcomes) %>%
         mutate(team='dire')
    y = rbind(yh, ya)
  }

  d = merge(y, x, by=c('game', 'team'), all=T, sort=F)

  ## the end

  return(d)
}



## wins
dw  = create.wide.for.apm(dp, dt, 'game', 'player',
                          outcomes='w', off.def=F, rep=F)
dwr = create.wide.for.apm(dr, dt, 'game', 'player',
                          outcomes='w', off.def=F, rep=T)
```

```
## gpm
dg  = create.wide.for.apm(dp, dt, 'game', 'player',
                          outcomes=c('gpm','k', 'xpm'),
                          off.def=T, rep=F)
dgr = create.wide.for.apm(dr, dt, 'game', 'player',
                          outcomes=c('gpm','k', 'xpm'),
                          off.def=T, rep=T)

## save
filename.w  = 'data/apm.wins.no.rep.rds'
filename.wr = 'data/apm.wins.with.rep.rds'
filename.g  = 'data/apm.gpm.no.rep.rds'
filename.gr = 'data/apm.gpm.with.rep.rd'
saveRDS(dw , file=filename.w )
saveRDS(dwr, file=filename.wr)
saveRDS(dg , file=filename.g )
saveRDS(dgr, file=filename.gr)
```

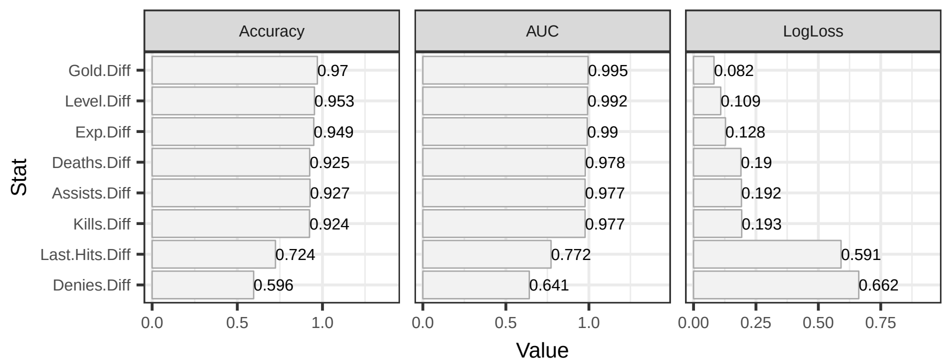

Figure 2: Metrics describing the strength of the relationship between various team statistics and winning. Note that Gold Differential is most strongly related to winning according to all three metrics.

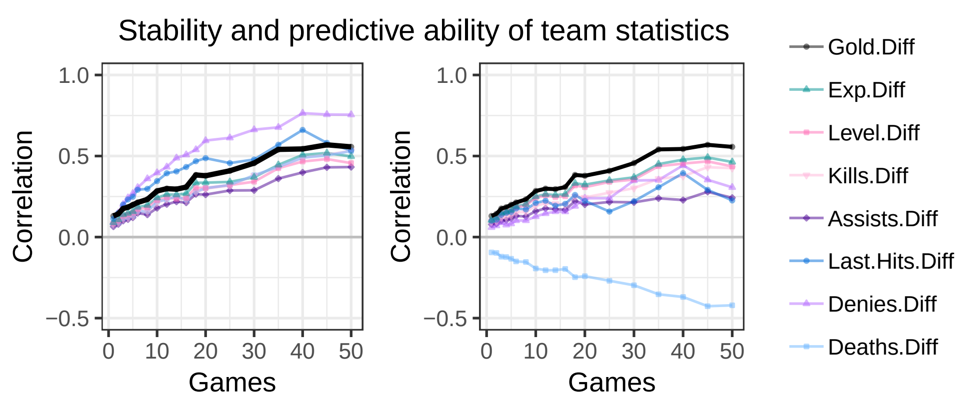

Figure 3: Stability of (left) and predictive ability of (right) team statistics for different numbers of games. The best predictor of Gold Differential is Gold Differential itself (black line on right).

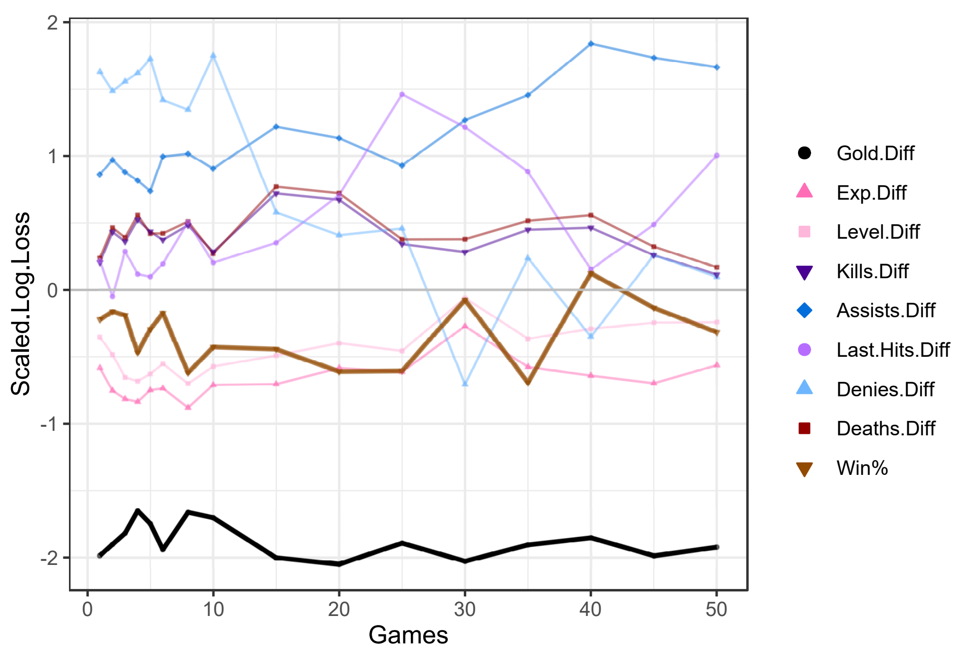

Figure 4: Log Loss of logistic regression models using team statistics in n games as the predictor, and Win% in a different n games as the outcome, for different values of n. Gold Differential (black) has the lowest Log Loss, indicating that it is the best predictor of Win% in out-of-sample games. In particular, Gold Differential is better than Win% itself at predicting out-of-sample Win%. For each value of n, the Log Loss has been centered and scaled so that all lines and values of n can fit easily on the same figure.

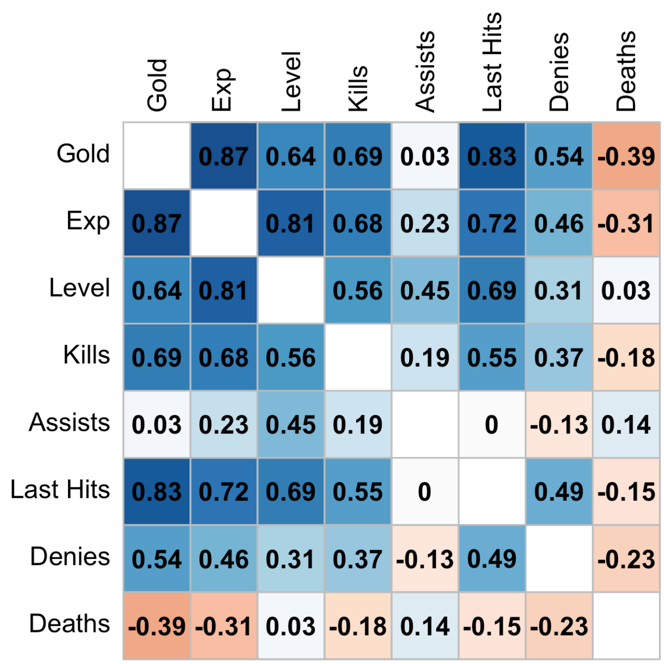

Figure 5: Correlations between player statistics.

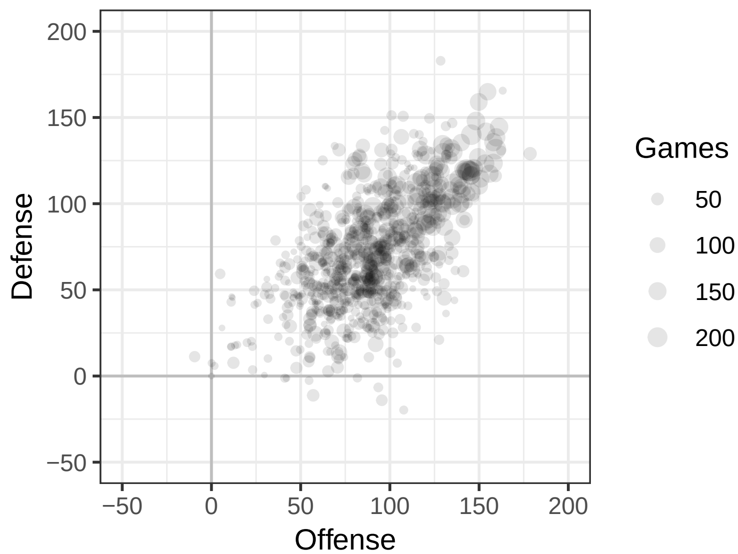

Figure 6: Posterior means for offensive vs defensive coefficients for players

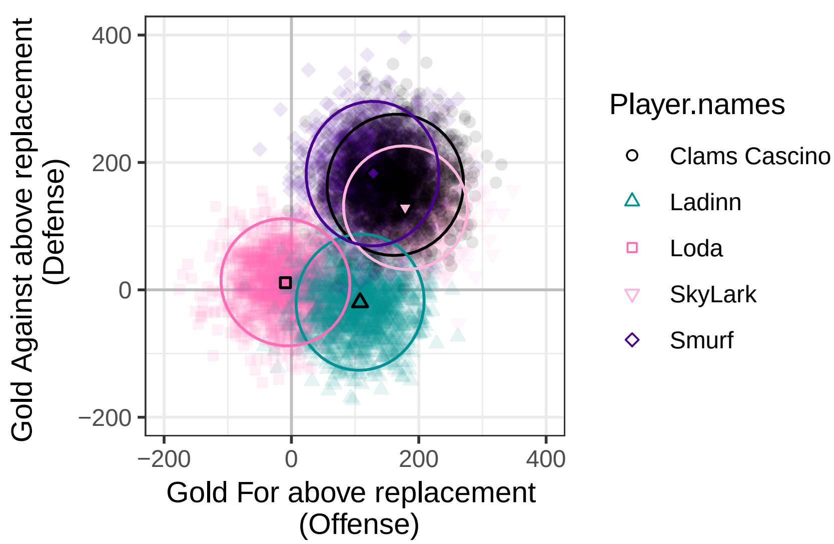

Figure 7: Example Posterior distributions for offensive and defensive coefficients for four players.

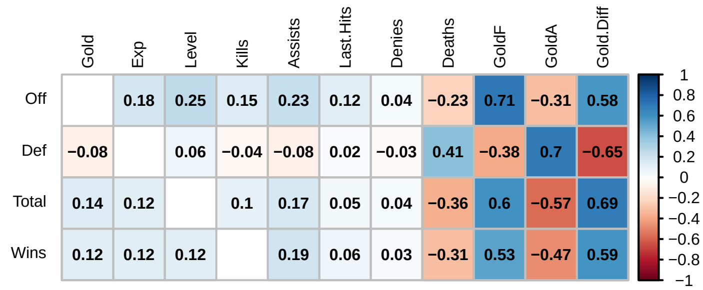

Figure 8: Correlations between offensive (Off) and defensive (Def) coefficients from the Gold model, off minus def (Tot), the coefficient from the Wins model (Wins), and various box score statistics.

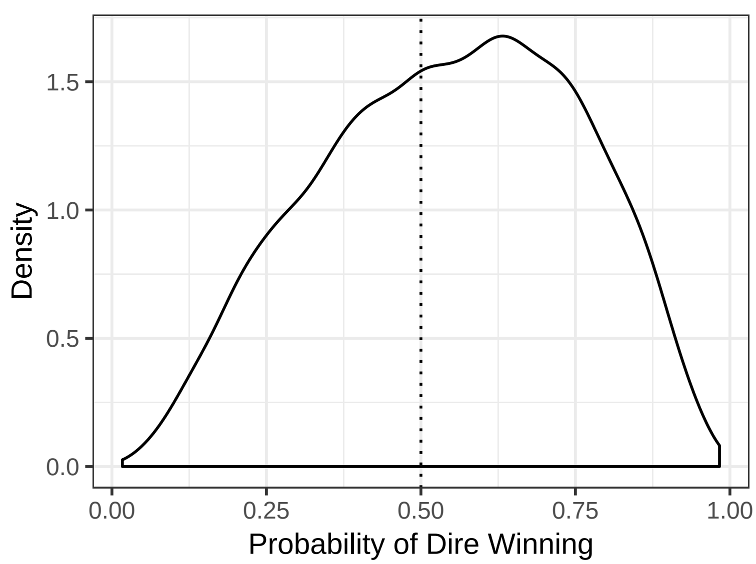

Figure 9: Probability of Tigers Beating Natus Vincere with Tigers Playing as Radiant

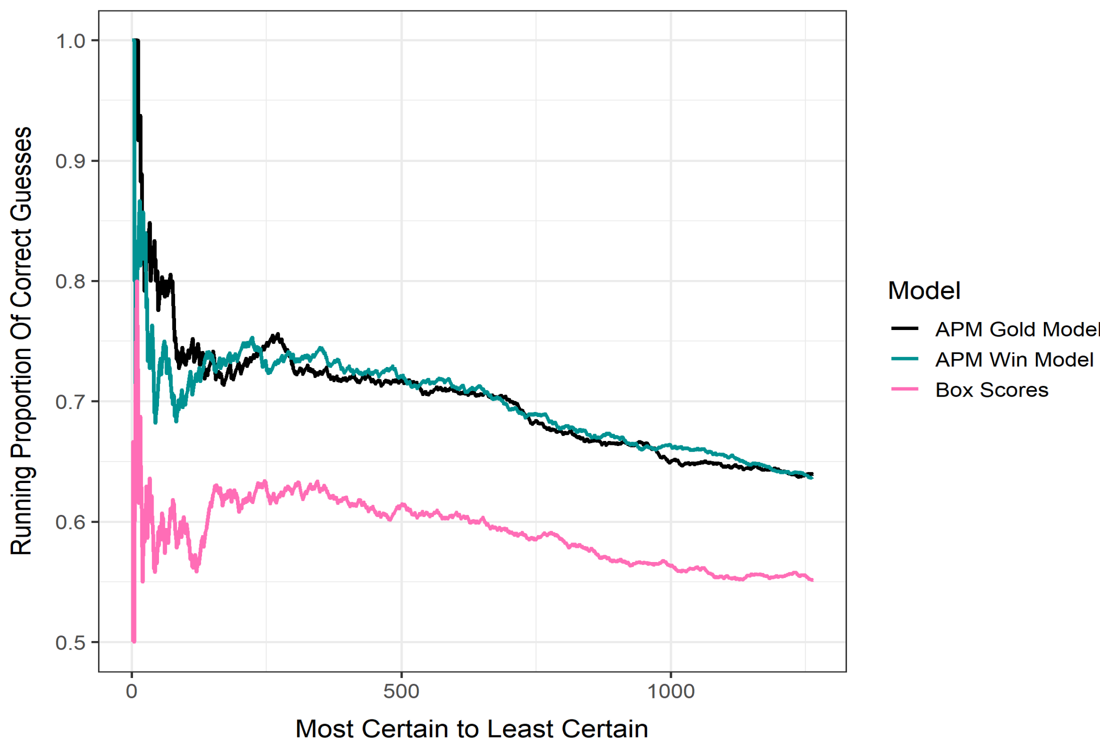

Figure 10: Proportion of correct predictions using models (15) and (2) compared to using a model based solely on box score statistics. As we move from left to right on the x axis we start with the games that we are most certain on (highest Δ Gold or highest win probability) and move to games with lower Δ gold or win probability. In general both models outperform the boxscore statistics consistently.