

BRM Win Rate Estimation based on DWG's Dominance in S10 World Championship for League of Legends*

Result based on estimated win rates calculated with Bayesian Regression Model

Zhendong Zhang

22 December 2020

Abstract

In year 2020, which is the Season 10 for League of Legends. DWG showed everyone their dominance against the world, and this is another LCK Champion since 2 years ago. Before now, the public thought LCK is becoming out-dated along with the game's version updates, but now DWG proved them wrong. Just as the former champions, DWG gained its champion with absolutely advantages among other teams. In the report, I will use Bayesian Regression Model for factors influencing the outcome of a game, exploring along the experience from DWG way to champion, and building a model to predict the performance of a team in World Championship.

Contents

1	Introduction	2
2	Data	2
2.1	LCK Data	3
2.2	World Championship Data	6
2.3	LCK & World Championship Data Comparison	8
3	Model	10
3.1	Variables Explanations	10
3.2	Model Principle	11
4	Results	11
5	Discussion	12
5.1	Strength	13
5.2	Weakness	14
5.3	Next Steps	14
6	Reference	14
7	Appendix	15

keywords: "League of Legends, DWG, S10 Championship, Bayesian Regression Model"

*Code and data are available at: <https://github.com/Craymate/lol>

1 Introduction

This paper will analyze why DWG (DAMWON Gaming) can dominant every game in Season 10 which just ended a month ago. DWG made its Champion with only 2 losses in championship, which represent the highest competitive level around the world. This paper will introduce a new tool for measurement of a team's performance by the estimated win rates. This is based on the team's match history from Season 10 (2020). League of Legends is a 5 vs 5 competitive game that are splited to 2 sides. Players have to gain gold and level advantages in order to win a game. The most large source of gold is from neutral resources and kills. Each opponent player kill could bring 300 golds, which could contribute a player's advantages. And the source of experience could also be kills, as well as minions and neutral resources kills. Gaining experience could make player level up, as most game do. The higher the level, the stronger a player.

The analysis are done on software **RStudio** [RStudio Team, 2019], based on language **R** [R Core Team, 2020]. In this paper, it will sufficiently use packages **tidyverse** [Wickham et al., 2019], **gridExtra** [Baptiste Auguie, 2017], with package **bookdown** [Yihui Xie, 2020] in order to produce output. And **readr** [Hadley Wickham and Jim Hester, 2020] in data cleaning.

For a whole team's performance, there are several factors need to be considered. Each game in within a map, and there are neutral resources that players could take. These neutral resources are excludable, which means if the one side took the resources, then there is nothing left for the other side. Then it is important to gain advantages with these resources, and this includes: "Herald", "Dragon", "Baron". In this paper, the team which took the first herald could be an important factor in the rival, thus it is a factor for teams' performance. Then for "Baron", unlike "Herald", this resource appears in the late game, so it wouldn't be important whether which team get it first. Then the total number of "Barons" taken is more important as a factor. "Dragon" is more common resources among the others, it appears more frequently, so "number of dragons as advantages of a team" could specifically reflect a team's competitive ability and advantages during games.

Firstly, it will discuss the variables in the LCK data, which DWG stays its dominant position on these variables, and trying to analyze the logic behind this. This is reasonable since these variables could increase the credibility of the model, by assuming the prior distribution from the analysis. Secondly, there will be variables with the same such dominance collected from WCS data, trying to analyse the differences of DWG's advantages between LCK and WCS. Furthermore, it will compare some variables for a same team between its performance in LCK and WCS. This would also be helpful to determine what changes could a more stressful and intensive environment bring to players and a whole team. These analysis would help us to build the model with more appropriate variables, with more wisely measures.

In this paper, it will use the average value for each variable across the team's history, and using the data from their own league "LCK" (League of Legends Champions Korea). These average values are being weighted by time, since the earlier data could reflect less for World Championship. The dependent variable would be explicitly the win rate by the team in world championship, which could be a representative rank for teams' performance. By aggregating these factors, it could conclude an estimated win rates for a whole team by Bayesian regression model which could fit a appropriate distribution for the data and a more precise model. In this way, this paper will explore the which factor is the most important to dominant a game from DWG's win rate, and it could be taken into consideration of the estimation model. In the future, it could predict the performance for a team in Season 11 (2021) by this model as well.

2 Data

The data is provided by Tim Sevenhuysen of OraclesElixir.com. It is free of charge, and is intended to be used in multiple ways, such as analysis, commentators and also fan-made products [Oracle's Elixir, 2015]. The data consists of 2 parts, one is the data from their local league LCK, and the other is from the world championship which is called WCS. This data provides an insightful view on each match's statistics, by offering gold differences at different time period, basic data such as kills, deaths, assists for each player, and neutral resources statistics. These information are in form of each match, with data corresponding to both winner and loser team. Of course the results of each match between 2 teams are also included, and this is

attribute to the exact time of the match began with precision to seconds. For league LCK, it is divided into 2 splits, which is spring split and summer split. The world championship happens after the playoff, while the playoff is after the summer split, entering by the rank points of each team. The time span for this whole process is nearly one year. Since this is an eSport competition, the game updates over time for reasons such as balance issue, bug issue and addition of contents. Thus there are different versions of game during this process, and this is recorded in the data.

Another key factor is on statistics level. From the data of every game, its statistics could reflect which team is the winner of a game. The winner team usually has the higher “Damage per minute”, since they could win most of the fights, they need more damages. Similarly, winner team could have more “Damage taken per minute”, but this depends on several factors, since the loser team could also have higher damage taken as they lose most of the fights. Then, a important index between these two statistics is “Kills per Minute”, it can explicitly reflect the team’s condition, because winner teams always have more kills eventually. In the duration of a game, the time at 15 mins is significant since the game is changing from early stage to mid stage at this. If a team can gain large advantages at or before this time, then it is most likely this team will win this game, as the late stage only depends the advantages a team accumulated earlier. At last, the “Gold difference at 15mins” and “Experience difference at 15mins” could show a team’s ability to get advantages in early-mid game, which is most important in winning a game.

The data uses amounts of dummy variables in recording whether the team gets the first mid turrent, first herald, first baron and first dragon, such as 1 for positive, 0 for negative. And these information are team level data, which would appear as NA (Not available) or simply blank under players’ statistics, and only presents under team’s statistics. There are also a kind of data including creep score difference at 15mins, gold difference at 15mins. For this kind of data, they are symmetric, which means that if the gold difference for team A at 15mins is 100, then it is -100 for its opponent team B. And the dummy variable mentioned before is also an example. Besides team level, it appears in players level, or more precisely in lane level. The 5 players of each team have their different roles, but the structures for teams are same, so one player could be compared with the player that has the same role in opponent team. There are statistics containing creep score differences at 15mins etc, and this is also symmetric.

The data has a small portion of missing data, which is recorded as partial in the data set, while others are complete. These partial data usually miss some of the statistics about a match, but there are still information in relevant of the missing part. This relevant data could be used as a alternative data for the missing one. Most of the partial data appear in the league LPL, which is the local league in China, since their league is owned by another corporations. In this report, the main focus is on LCK data, and there is only 1 match has partial information. This is the match happened in summer split of LCK at 2020-08-28 04:56:26, with Gen.G vs. Afreeca Freecs.(2 other teams from LCK) This match is lack of gold difference, creep score difference, experience difference such comparative statistics, and this is irreplaceable. Thus it is removed in cleaning process.

In the data cleaning procedures, the team and player levels of data are averaged by 5 in each match with respect to their level. By dividing team and player level data, the average is performed for each of their variables, and aggregating them back in the end. At the last, the data contains 2 rows for each match, one is for team A, and the other is team B in the match. Also, as the time has influence in the performance, since the game updates could make itself completely change from spring split to world championship. The players’ performance could be changed, as well as teams. For example, a player who showed dominant performance in spring split could be totally disappointed in WCS. Thus each calculation of averages is weighted by time difference between the time of the match and the beginning time of WCS, in order to increasing the accuracy of the model of estimation.

2.1 LCK Data

As highlighted in figure 1, the S10 champion’s win rate in their own league is the third place overall, which falls behind Gen.G and T1. As we can see in table 1, DWG has the win rate of 67.7%, while T1 and Gen.G have 67.7% and 69.8% respectively. Since DWG also won the champion in LCK, the reason its win rate is the third is because of T1 and Gen.G have more wins. As in table 1’s data, T1 and Gen.G have 65 and 67

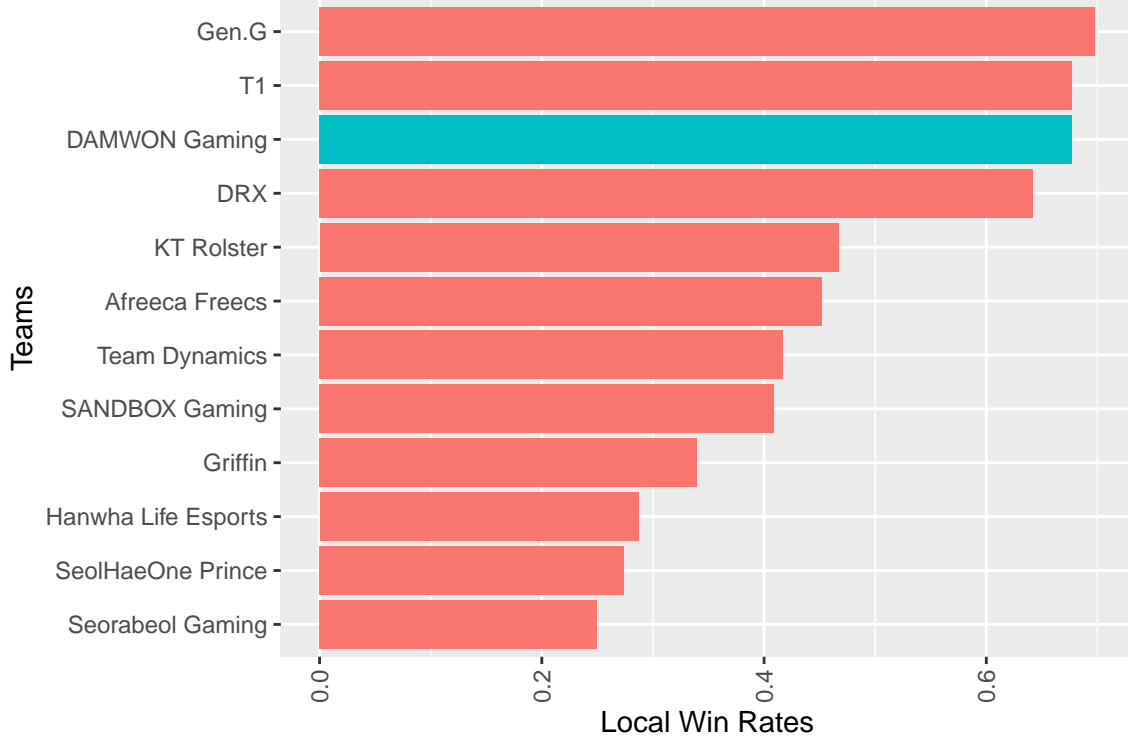


Figure 1: Winrates for Teams in league LCK

wins respectively, while DWG has 63 wins. It is due to T1 and Gen.G played more games than DWG, thus there are more wins.

More wins could lead to a higher win rate since the equation of win rate is:

$$\text{Win Rate} = \frac{\text{Total Number of Wins}}{\text{Total Games Played}}$$

As both the numerator and denominator increase, the win rate also increases under certain conditions.

Kills worth more than the gold which a team would receive. It also contains the time cost, when opponents die, they have reviving time, which means they can't access or receive any resources during the time. In this period of time, the alive team could get large amount of neutral resources without interference such as Baron, Dragon and Herald. As a strong team always has higher kills per minute, they could have higher chance to win since they have more opportunities. As shown in figure 2 and table 2, DWG gains the highest KPM, with 0.46 kills per minute. This means if a team is DWG's opponent, then on average, this team will lose one player to gain golds and numerical superiority in about 2.17 minutes. The equation is as following:

$$\text{Time Needed to Gain Advantages} = \frac{1}{KPM}$$

The second is Gen.G, which is 0.449. Although Gen.G has more games played, KPM is a fair measurement regardless the number of games played. Gen.G could gain its advantages on average 2.23 minutes. This could reflect DWG's dominance in some aspects. The lowest KPM is from Griffin, which is 0.237, means they need 4.22 minutes to gain advantages. After snowballing effect, this gap could be huge.

Gold difference could be the most intuitive measure to reflect the situation of a game. With more gold, players could upgrade their equipment more, thus could bring them more powerful effectiveness for combat. Since snowballing effect exists in this game, the advantage in early game could be significant in winning a game. Most of the large neutral resources are available in early stage, then the team with more gold has

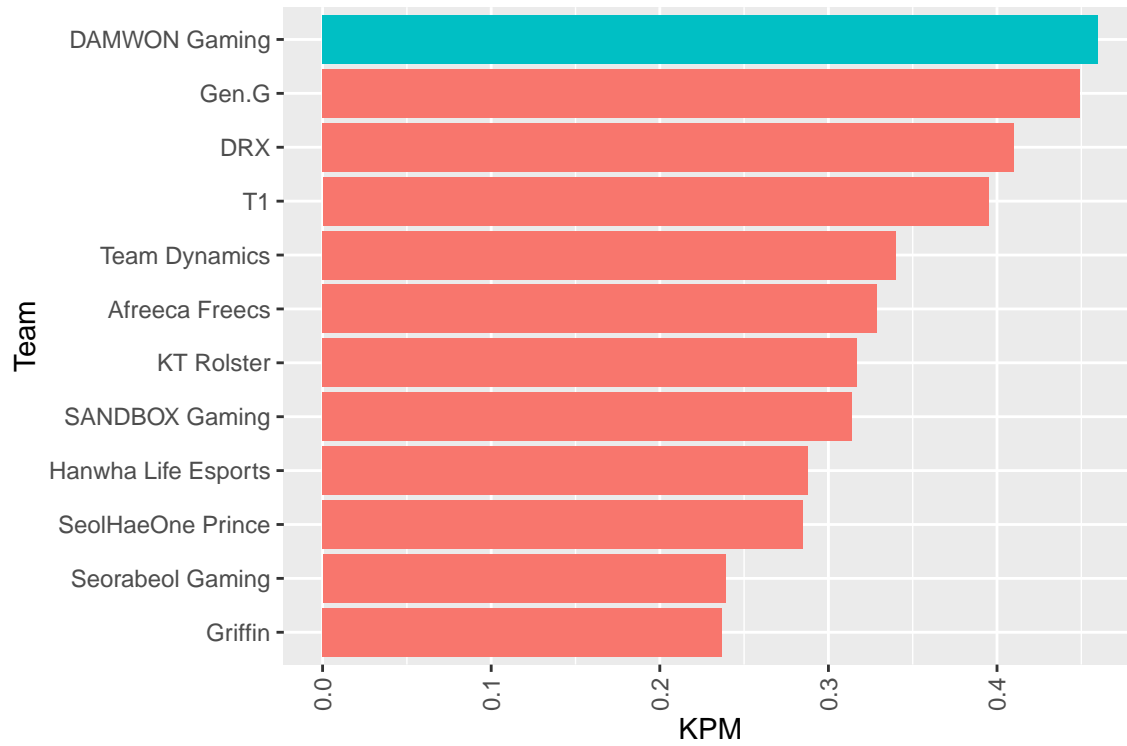


Figure 2: KPM for Teams in league LCK

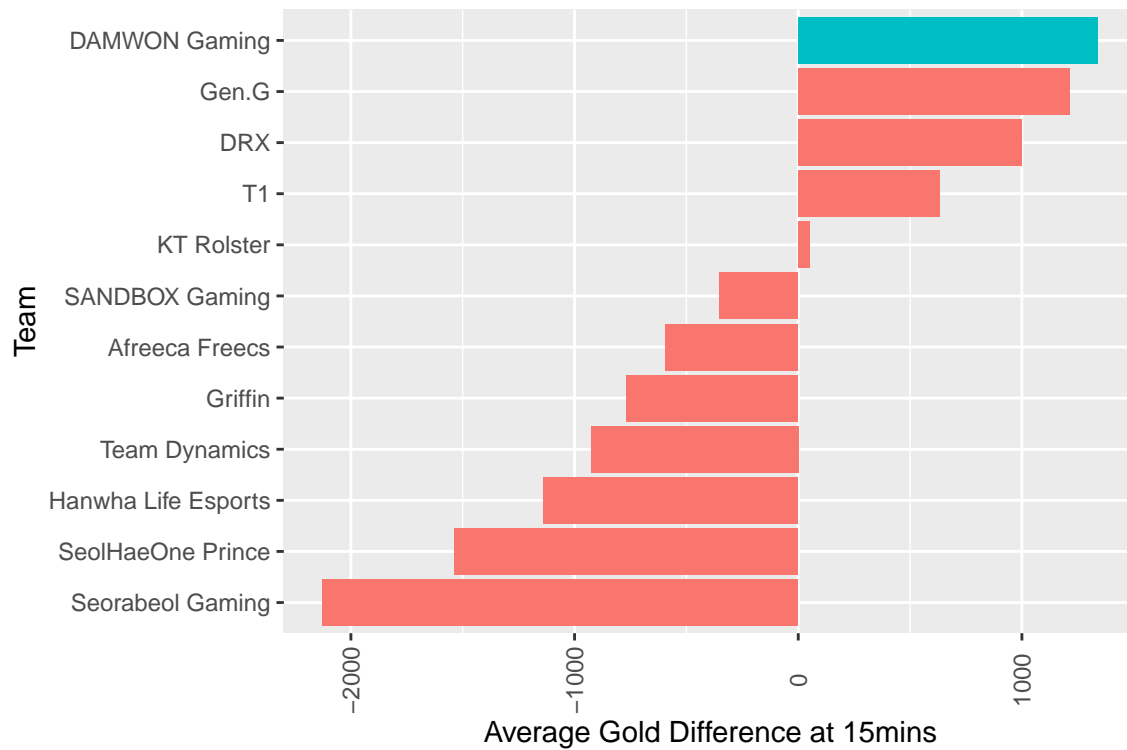


Figure 3: 15mins Gold Difference for Teams in league LCK

higher probabilities to take control of these. As shown in figure 3 and table 3, DWG still dominates this statistics, with 1341.022 gold advantages at 15mins, ranked at the first among all other LCK teams. Note this measure is an average value, meaning they could gain 1341.022 gold advantages on average for every game, and this is even at 15mins. This rank conform to the rank of KPM, as the kills could bring them gold lead.

2.2 World Championship Data

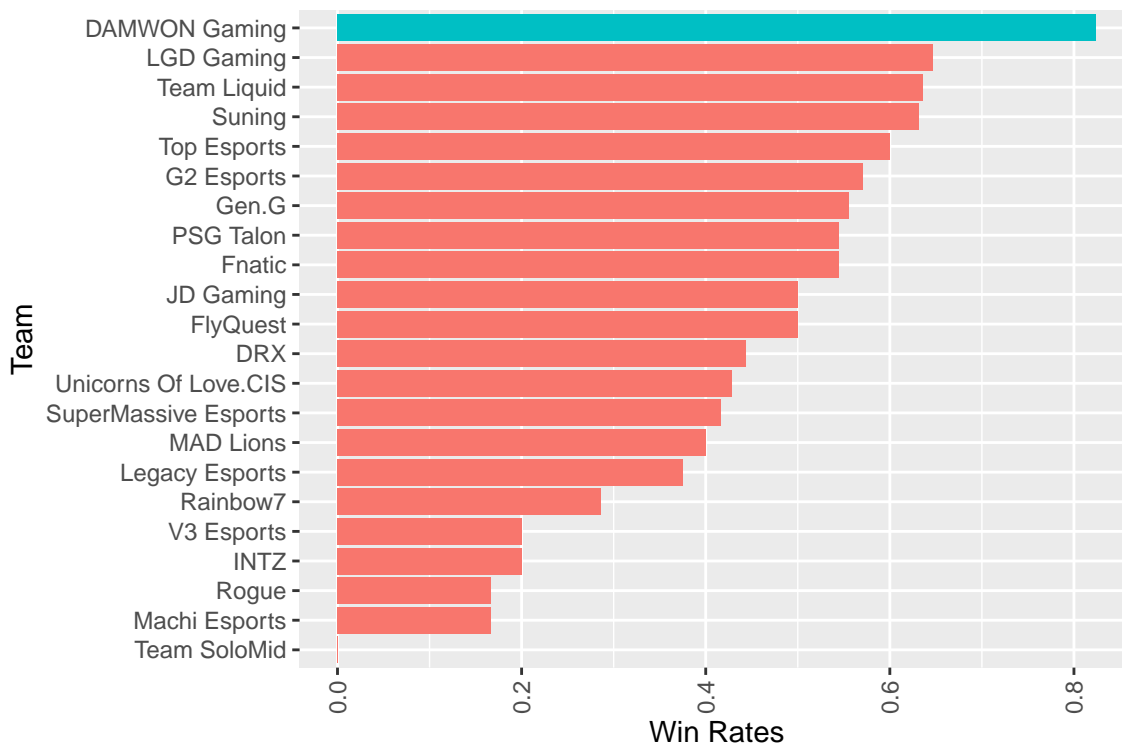


Figure 4: Win Rates for Teams in World Championship in league LCK

The win rate in WCS won't exactly reflect the rank in the competition. As shown in figure 4, although DWG remains the first (as highlighted), the WCS second-place Suning is the forth in rank of win rate. For teams like LGD Gaming and Team Liquid, they have more games played. Same as discussed in LCK data part of win rate, more games could lead teams with higher win rate. These additional games were played for a stage called "Play-in Stage". Each local league's last place will go to this stage in order to gain qualification for WCS. In cleaning process, the data was capped to matches after 2020-09-25, since that's the time of the main event. But for the situation here, it is because in BO3 or BO5 (Best of), the teams won with 2-0 (2 wins 0 loss) or 3-0 (2 wins 0 loss) in BO3 and BO5 respectively, could make more contribute into win rate. Since DWG only loss 2 games in WCS, this is why DWG could remain its first place.

In figure 5, the significant lead of DWG in average number of dragons advantages could be the key point in winning the WCS. AS shown in table 4, DWG could take 2.118 more dragons on average, while the second, PSG Talon, could only take 0.727 more dragons. This is the thing that DWG has huge lead among other teams, which mean other teams don't want to or can't compete for dragons against DWG.

In WCS, as shown is figure 6, DWG keeps its advantage in LCK, which is to take gold advantages at 15mins. DWG couldn't make a big gap between the second on this, since it is another LCK team, DRX, which DWG didn't take much advantages on gold in LCK. But unlike DRX, Gen.G, which ranked second in LCK's gold advantages chart and is close to DWG, didn't make its performance in WCS. As shown in table 5, DWG's advantages is 1939,647 golds, and DRX is 1876.333 golds, while Gen.G drops to 896.556 Golds.

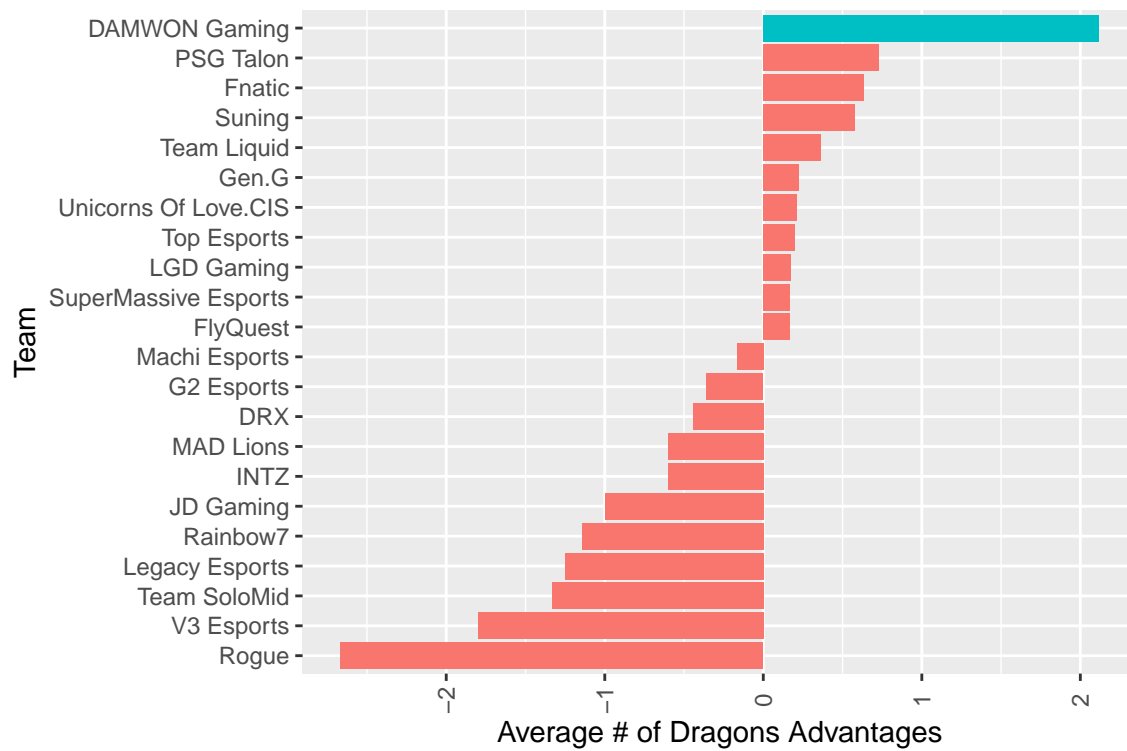


Figure 5: Average Number of Dragons Advantages for Teams in World Championship

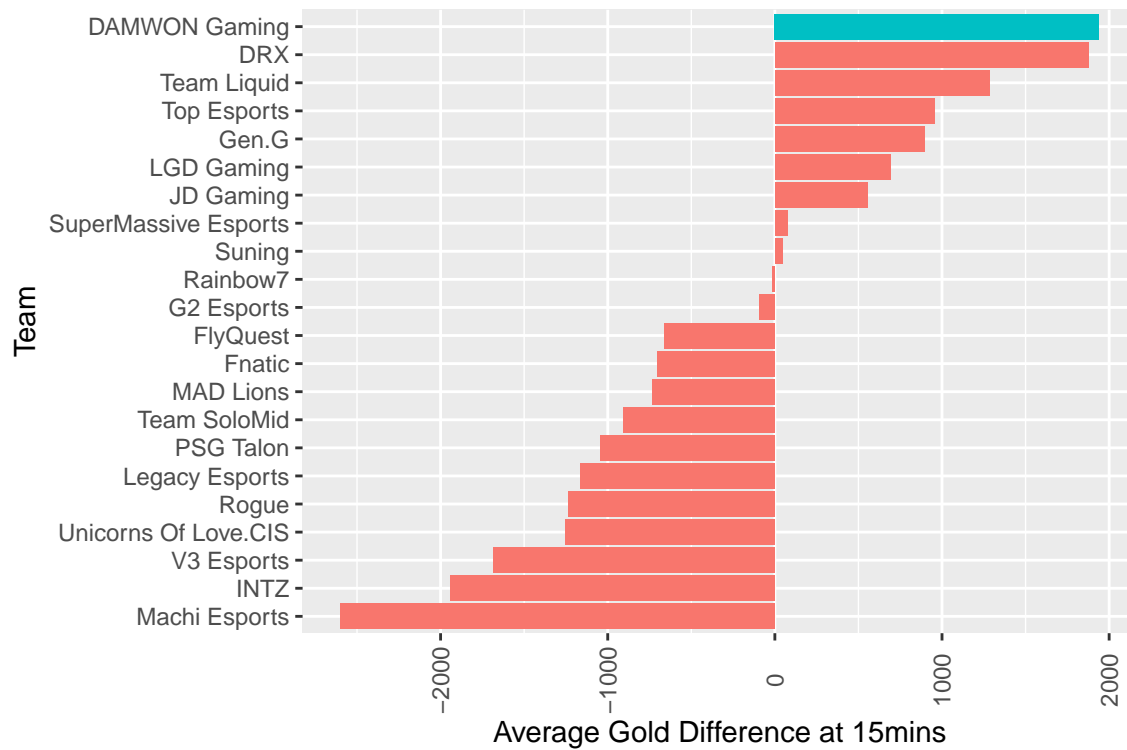


Figure 6: Gold Difference at 15mins for Teams in World Championship

2.3 LCK & World Championship Data Comparison

In this section, the main focus is to determine whether World Championship could bring down teams' performance, as well as players' performance. Since WCS has more audiences and is more valuable competition, players will face the external pressure from multiple aspects, such as social media, economics and glories. Thus players could be more stressful than they are in LCK which is the local competition. Also, the opponent teams could be much stronger, as they all come from the top of different leagues. These could make players to performance poorer than usual, and it will influence the model as well, since the model is about to use local statistics to forecast the world performance. So examining and comparing these data could be useful in increasing the credibility of the model.

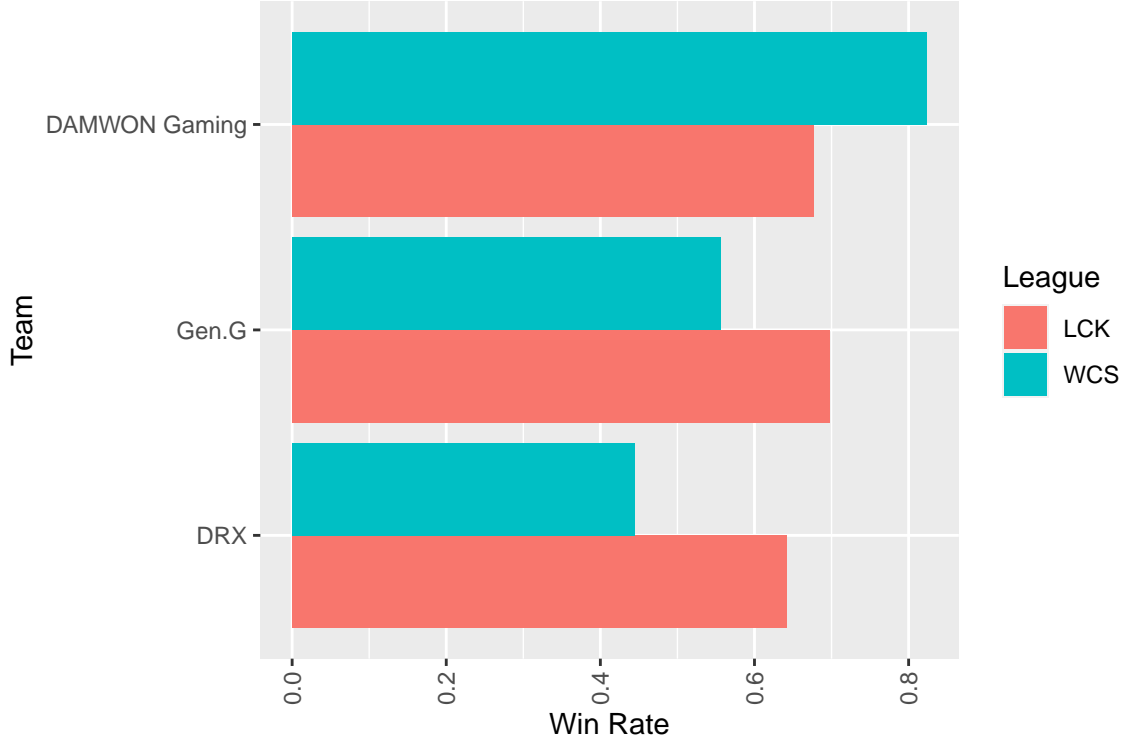


Figure 7: Comparison for Win Rates between LCK and Championship

As presented by figure 7, both Gen.G and DRX has poorer win rate comparing to their win rates in LCK. But unlikely, DWG makes improvement on win rate instead. Since DWG could claim its champion, it is normal to have the highest win rate, but they made their one step forward as well. Based on the number from table 6, DWG's win rate increased with 14.7% in WCS, comparing with LCK. While Gen.G and DRX decrease 14.2% and 19.8% respectively.

There is a similar pattern as win rate, DWG's kills per minute increases in WCS, while the other 2 decrease, as shown in figure 8. Calculating based on table 7, the increase of KPM for DWG is 0.11, meaning DWG's average time to gain kills decreases by 19.298%. While DRX and Gen.G take 14.846% and 10.591% more time to gain a kill, respectively. The percentage changes is based on equation:

$$\Delta Time\ to\ Gain\ Kills = \left(\frac{1}{KPM_{WCS}} - \frac{1}{KPM_{LCK}} \right) \times KPM_{LCK} \times 100$$

Unlike before, DRX made its steps with DWG, which gain progress in WCS on gold advantages at 15mins. Gen.G still shows decreases in their performance in WCS. But DWG could keep its advantages on every aspects, and this is horrible in WCS, since only a few teams could have this achievement. From figure 7, 8, 9, we can identify that DWG has its steady increasing performance during the WCS, while other teams do not, which means that is the key to claim the champion.

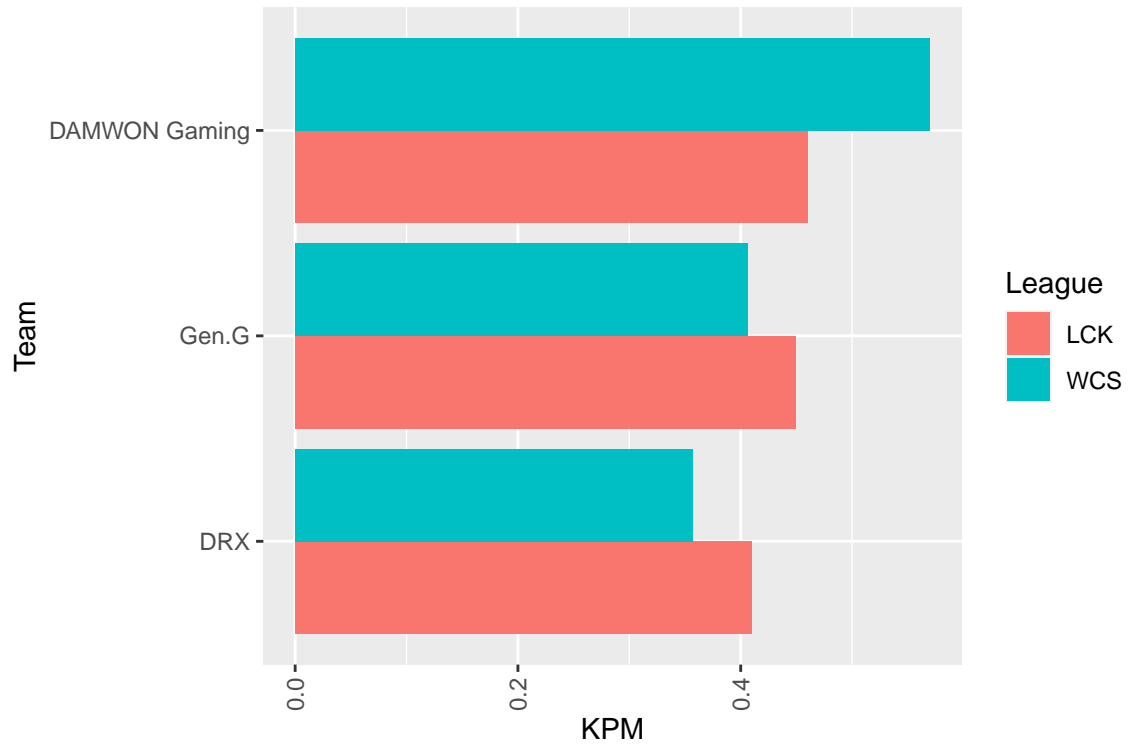


Figure 8: Comparison for KPM between LCK and Championship

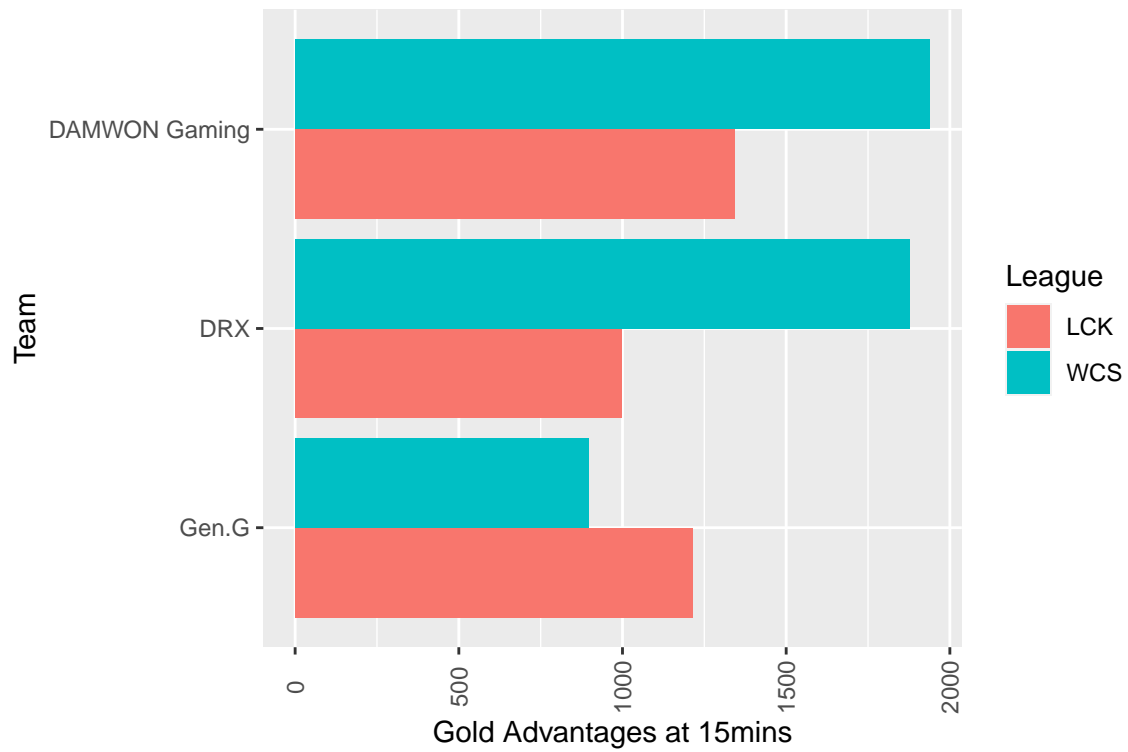


Figure 9: Comparison for 15mins Gold Difference between LCK and Championship

3 Model

The model is built based on Bayesian Regression Model. The dependent variable, y , is the win rate of each team in WCS as introduced before. The independent variables are each team's local win rate, average KDA on team level, vision score per minute, creep score per minute, damage share, team's average dragons advantages, first herald rate, first tower rate, number of barons taken, team level kills per minute, damage taken per minute, damage per minute, gold advantages at 15mins and experiences advantages at 15mins.

The KDA (Kill Death Assist Ratio) could be a measurement for a player's performance in each game, then it could get a expectation of a player's KDA in the World Championship. The KDA is calculate by the ratio between number of kills, number of deaths and the number of assists. Increasing in number of deaths could lower the KDA, while increasing in the other two would make KDA higher. Also, vision score per minute and creep score per minute could reflect a player's ability to sense the macro of the game and ability to get the steady source of gold. The map contains a number of unseen area, which is called the war mist. Players can put a ward in a certain area to make the area visible for a period of time, and this is costly. So whether a player can use their ward to detect enemy is the way to determine player's sense of macro, the vision score is the measurement of it. If a ward doesn't detect any enemy, then the vision score wouldn't increase. The creep score is referred to another neutral resources in the game, which is minion. Minions appear frequently and periodically, as each of them worth 15 golds to 90 golds. They comes as a wave of 6 minions, and the waves come each 30 seconds. Ensuring stable gold gains from minions could make the game more stable in winning. This is what creep score does, and by calculating it as per minute could show a player's status in a game.

3.1 Variables Explanations

For KDA which is a measure in determining a team's performance. This is calculated by:

$$KDA = \frac{Kills + Assists}{Deaths}$$

This measurement increased when a team shows a better performance. Simply when deaths goes down, the KDA would be up, oppositely, KDA goes down when deaths goes up. Although in most games, the number of deaths for a whole team is impossible to be 0, since this is suggesting that this team is making no mistakes in the whole game. There is still small part of games with 0 deaths of a team. In this case, the KDA should goes to infinity since the denominator is 0. I handled this by adjusting its KDA to the highest KDA among the whole competition instead of infinity. In this way, it would be enough to reward this brilliant of 0 deaths for the team.

Snowballing effect is an important law in this eSport game. It means if a team could gain advantages, then the team could have higher probability to gain more advantages. Because advantages could make a team stronger, such as equipment, experience level and advantage often comes with more initiative, a team with advantages have more chances to win a game. But this is not absolute, sometimes advantages could be vanished by a good skilled player, or a better team cooperation. Of course, this is why eSports matches are exciting. Despite of this, snowballing effect is still an significant point, thus a team with advantages in early stage is more likely to win the game.

For first herald rate and first tower rate, as introduced, they are the dummy variables in reflecting whether a team could claim its early stage advantages in order for snowballing, with outcomes of 1 (Positive) and 0 (Negative). These 2 factors are calculated based on:

$$Rate = \frac{Number\ of\ Postives}{Number\ of\ Postives + Number\ of\ Negatives}$$

Same for gold advantages at 15mins and experiences advantages at 15mins, they are counted for the early game advantages as well.

For the rest variables, they are presented in order to reflect what a whole game's trend be. For example, the winner teams usually have a positive number of dragons as advantages, and a higher KPM, Vision Score per minute and Creep Score per minute, etc. Thus these all are possibly influencers of the win rate at world.

3.2 Model Principle

The Bayesian Regression Model is based on Bayesian Statistics which has the equation as:

$$P(\Theta|data) = \frac{P(data|\Theta) \times P(\Theta)}{P(data)}$$

Θ represents the vector of variables that we are interested in. It assume a Prior Distribution which is $P(\Theta)$ in the RHS, which means these variables follow a certain distribution. And in the LHS, the thing we calculate is called the Posterior Distribution, which is the distribution of the interested variables based on our data. In this way, we could conclude the relationship between our data and the variables of interests, with the prior distribution assumed based on the analysis from Data section.

From the *brm* [Paul-Christian Bürkner, 2018] function in **brms** package [Paul-Christian Bürkner, 2017], it iterates to get the posterior distribution by a method called “Gibbs Sampling” with MCMC (Markov Chain Monte Carlo). The model runs based on R, by the software RStudio.

Assume we have k interested variables, thus there are $k+1$ parameters including intercept, the equation is:

$$Y = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_k \times x_k$$

The steps are as following:

1. Setting up a initial value for each parameters in $\Theta : \beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)} \dots \beta_k^{(0)}$
2. Using samples from each Marginal Posterior Distribution to update:

$$\begin{aligned} &P(\beta_0^{(m+1)}|data, \beta_1^{(m)}, \beta_2^{(m)}, \beta_3^{(m)} \dots \beta_k^{(m)}) \\ &P(\beta_1^{(m+1)}|data, \beta_0^{(m+1)}, \beta_2^{(m)}, \beta_3^{(m)} \dots \beta_k^{(m)}) \\ &P(\beta_2^{(m+1)}|data, \beta_0^{(m+1)}, \beta_1^{(m+1)}, \beta_3^{(0)} \dots \beta_k^{(0)}) \\ &\vdots \\ &P(\beta_k^{(m+1)}|data, \beta_0^{(m+1)}, \beta_1^{(m+1)}, \beta_2^{(m+1)} \dots \beta_{k-1}^{(m+1)}) \end{aligned}$$

3. Repeating step 2 to update $\beta_i^{(m)}$

The string of values produced by these 3 steps is called Markov Chain, with expectation that it could converge to an equilibrium distribution in the end. This equilibrium is called Joint Posterior Distribution which is our model. The function *brm* [Paul-Christian Bürkner, 2018] takes 4000 iterations and 4 Markov Chains in order to get the result.

The output of the model is the estimated win rate which might be negative in some cases, but this could still reflect the rank outcome of each teams. Then we adjust every values to its 100 times in order to get percentages for comparing easily.

Since every variable in the data set is averaged for each match, but we need a single value for each variable among teams. Then we further averaged these data with respect to each team, and this is a weighted average.

4 Results

By re-applying the model onto LCK data, we could get estimated distribution of win rate in WCS.

As shown in table 8, the first 3 strongest teams are the same as before. Teams follow them are T1 and Seorabeol Gaming etc. Team T1 is estimated to have 13.5% to win its champion, and Seorabeol Gaming has 4.73% chances. In order to verify the estimation, we have to filter the result with only DWG, Gen.G and DRX left, since they are the only 3 LCK teams in WCS and with the actual win rates.

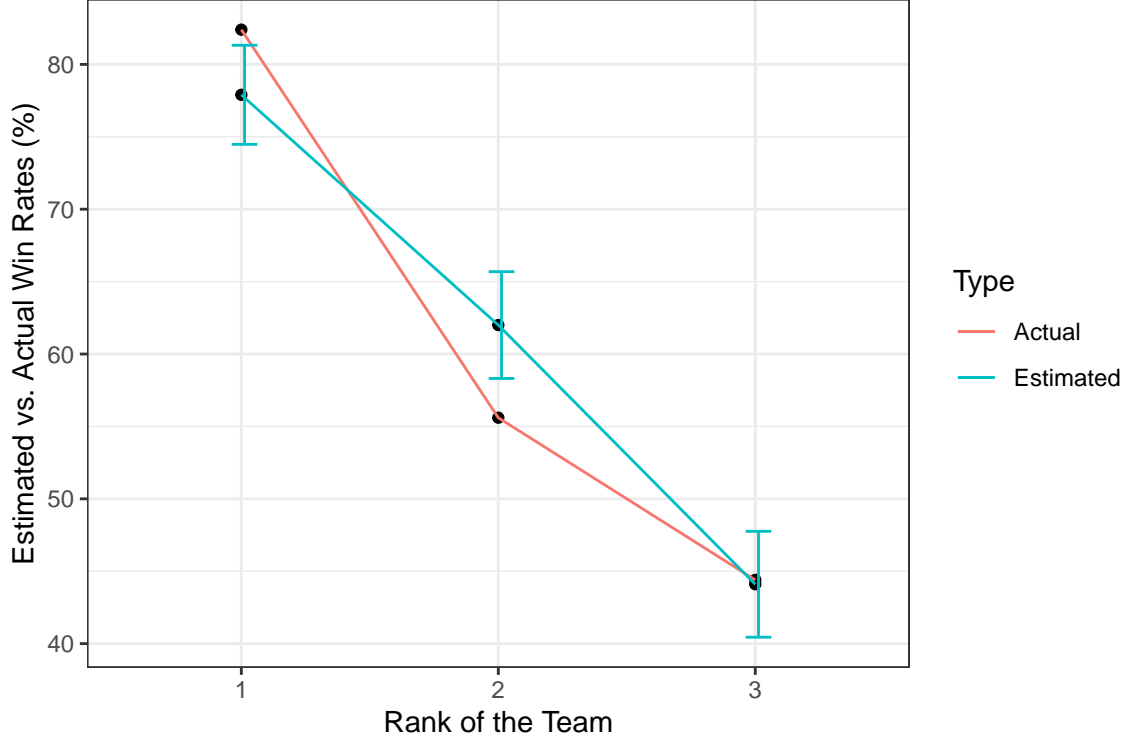


Figure 10: Estimation Errors' Trend

For LCK teams, based on figure 10, there are over-representations for the second and third places, and under-representation for the first place. However, these estimates don't deviate much from the actual data, and they are all within the standard deviation.

From figure 11, under the estimation, half of LCK teams would have 0% chance to win the world champion. There is a huge decrease between the third and forth rank, since the actual win rates for teams after the third place are 0 in training data set. Also, we can identify the deviation reaches the maximum at the forth rank, with decreasing to both tails. This is similar to normal distribution, so it means this is a good fit.

5 Discussion

Based on the data section, I assumed certain prior distributions for the variables of interests. And they are:

$$Local\ Win\ Rate \sim Normal(0.5, 0.5)$$

$$Gold\ Advantages\ at\ 15mins \sim Normal(0, 10000)$$

$$Experience\ Advantages\ at\ 15mins \sim Normal(0, 10000)$$

$$Number\ of\ Dragon\ Advantages \sim Normal(0, 6)$$

$$Kills\ per\ Minute \sim Normal(0.25, 2)$$

From the analysis before, only DWG could show its dominance in these variables, and the most teams are clustered at the middle part of the distribution, with some excessively low teams. Thus I decide to use Normal distribution on these variables. For win rates, most teams are clustered at 50% , with lower bound 0% and upper bound 100%. Gold Advantages and Experience Advantages at 15mins present identical distributions since they are the symmetric variables. Then by Normal distribution with expectation to be 0, and deviate by 10000, it could be clear expressed. Also due to experience, as usually seen in matches, 10000 golds and experiences advantages could bring a team to win in most cases. The maximum number of dragons in a

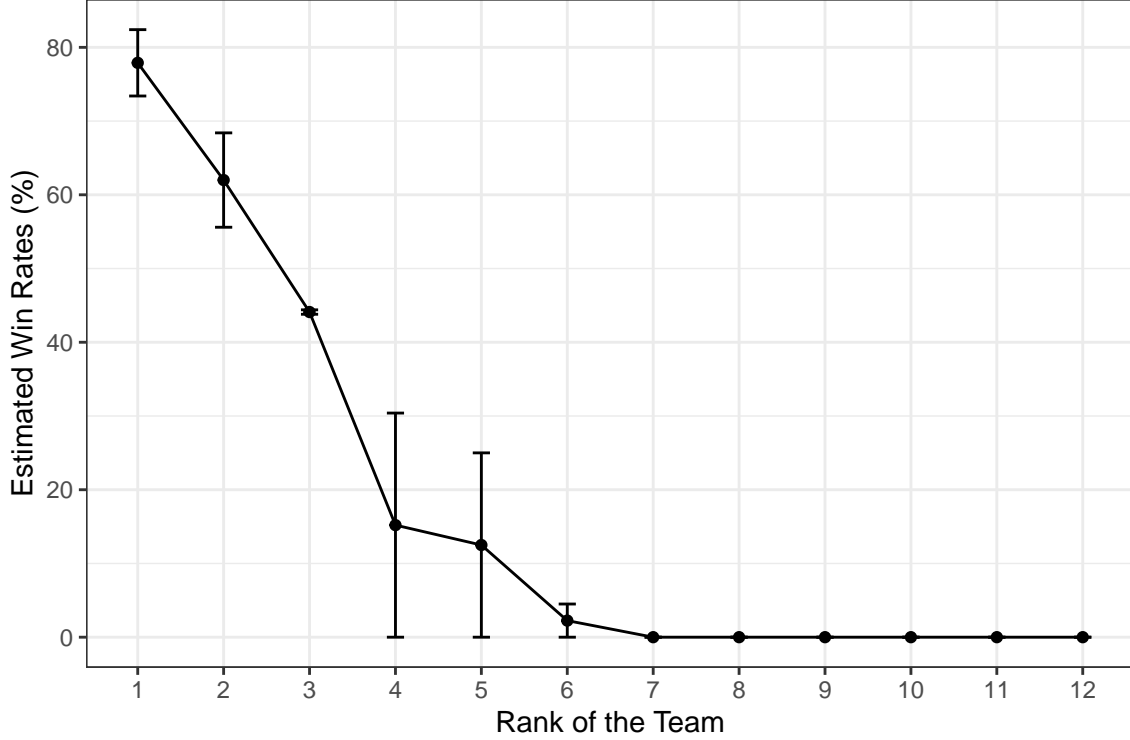


Figure 11: Estimation Win Rates' Trend

map is 7, thus with expectation to be 0, the standard deviation is 6 dragons in our data, following Normal distribution as well. Same for KPM, the mean in data set is 0.25, with deviation within 2 following Normal distribution, as introduced before.

We can identify from the estimation that there is a large gap on win rates between the third place and the forth. The probably explain the reason that most leagues pick the top 3 teams into WCS. With this year, which is Season 10, the official permitted leagues LPL and LEC to have 4 teams instead of 3, since they made their large achievements in last year. Although they have 1 more team, the results don't seem to be much different from before, because the additional teams got knocked out quickly. This conform to our estimation that there are large win rate gap between the first 3 teams and the rest.

5.1 Strength

The strength of this model is that I weighted each case by its time differences. Since the WCS starts at 2020-09-25, the teams' performances would change a lot along with time, as discussed in section 2.3 . Thus by implementing this equation, the training data is weighted by time. For each variable of interests used in the model:

$$Data = \{\alpha_1, \alpha_2, \alpha_3 \dots \alpha_k\} \quad \alpha_i = \text{Individual Match}$$

$$Variable \text{ Used in Model} = \sum_{i=1}^k \frac{Variable \ Value_i}{Time \ Differences_i}$$

Through this way, we could take account of the weakened in performance due to time. The closer match history from WCS, the more contributions made to variables value, which means a stronger effect. With the knowledge of prior, we could make use of this as advantages, since the training data set is still not large enough in order to build a high precision model, which is LCK match records in year 2020. By advantages

gained from assumed priors, we could make the training data sets more efficient by the probabilities of variables from distribution. It could offset the limitation of data in some extent, bringing us a more realistic estimation.

5.2 Weakness

However, this model is based on Bayesian inferences. When we setting the prior distributions, we could set them wrong, since the experiences won't always conclude the correct results. Also, our analysis are limited, and it influences our selection of variables, and also the prior distributions. Although Bayesian inference could offset the effects brought by small data sets, it can't eliminate all of the effect, so the trade-off between wrong assumptions and limited size effects should be an important consideration. Since there are only 3 teams could make into WCS, then for rest teams, we don't have their samples win rates in WCS. By assuming their win rates to be 0, it sacrifices their actual probability of winning a championship. This could also be the reason why there exists a gap on win rate for third and forth places, since it is due to the rule issues. Champion effects are an important factor in competitions. As the game updates regularly, the strong champions would vary along with version updates. And each player have the champions they are good at, so this could be an important factor on performances. But this model doesn't take this into account as well, due to limited data set.

5.3 Next Steps

In the future, I will further the analysis on the prior distributions, trying to find their deeper relationships. From this, I could balance the data size limitation issue and the wrong assumptions issue. Also, I will take account of more historical data such as 2019, 2018, etc. From this way, I could trace the develop pattern from different versions of game, and the key factor for a world champion in difference years. By then, it could make the prior assumption more reasonable and logic. Also, with more data, I could analyse the statistics for each player, in order to make the model more flexible. Thus we could analyze a team's performance by each player. Since a team's lineup would change every year, we could break the limitation that we could only use 1 year data in analysis for a team in this way.

6 Reference

- Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for “Grid” Graphics. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra>
- Hadley Wickham and Jim Hester (2020). readr: Read Rectangular Text Data. R package version 1.4.0. <https://CRAN.R-project.org/package=readr>
- Oracle's Elixir (2015). LCK and WCS Data Sources. URL <https://oracleselixir.com/>
- Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01
- Paul-Christian Bürkner (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10(1), 395-411. doi:10.32614/RJ-2018-017
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Yihui Xie (2016). bookdown: Authoring Books and Technical Documents with R Markdown. Chapman and Hall/CRC. ISBN 978-1138700109
- Yihui Xie (2020). bookdown: Authoring Books and Technical Documents with R Markdown. R package version 0.21.

7 Appendix

Table 1: LCK Wins and Win Rates Statistics

	Total # of Wins	Local Win Rates (%)
<i>DAMWON Gaming</i>	63	67.7
<i>T1</i>	65	67.7
<i>Gen.G</i>	67	69.8
<i>KT Rolster</i>	43	46.7
<i>SANDBOX Gaming</i>	38	40.9
<i>SeolHaeOne Prince</i>	23	27.4
<i>Afreeca Freecs</i>	42	45.2
<i>Griffin</i>	17	34
<i>Hanwha Life Esports</i>	25	28.7
<i>DRX</i>	68	64.2
<i>Team Dynamics</i>	20	41.7
<i>Seorabeol Gaming</i>	2	25

Table 2: LCK KPM Statistics

	Kills per Minute
<i>Afreeca Freecs</i>	0.329
<i>DAMWON Gaming</i>	0.46
<i>DRX</i>	0.41
<i>Gen.G</i>	0.449
<i>Griffin</i>	0.237
<i>Hanwha Life Esports</i>	0.288
<i>KT Rolster</i>	0.317
<i>SANDBOX Gaming</i>	0.314
<i>SeolHaeOne Prince</i>	0.285
<i>Seorabeol Gaming</i>	0.239
<i>T1</i>	0.395
<i>Team Dynamics</i>	0.34

Table 3: LCK Gold Advantages Statistics

	Gold Difference at 15mins
<i>Afreeca Freecs</i>	-593.022
<i>DAMWON Gaming</i>	1341.022
<i>DRX</i>	998.094
<i>Gen.G</i>	1214.811
<i>Griffin</i>	-768.458
<i>Hanwha Life Esports</i>	-1139.209
<i>KT Rolster</i>	53.511
<i>SANDBOX Gaming</i>	-353.196
<i>SeolHaeOne Prince</i>	-1537.096
<i>Seorabeol Gaming</i>	-2126.5
<i>T1</i>	633.326
<i>Team Dynamics</i>	-927.25

Table 4: WCS Dragons Advantages Statistics

	Average # of Dragons Advantages
<i>DAMWON Gaming</i>	2.118
<i>DRX</i>	-0.444
<i>FlyQuest</i>	0.167
<i>Fnatic</i>	0.636
<i>G2 Esports</i>	-0.357
<i>Gen.G</i>	0.222
<i>INTZ</i>	-0.6
<i>JD Gaming</i>	-1
<i>Legacy Esports</i>	-1.25
<i>LGD Gaming</i>	0.176
<i>Machi Esports</i>	-0.167
<i>MAD Lions</i>	-0.6
<i>PSG Talon</i>	0.727
<i>Rainbow7</i>	-1.143
<i>Rogue</i>	-2.667
<i>Suning</i>	0.579
<i>SuperMassive Esports</i>	0.167
<i>Team Liquid</i>	0.364
<i>Team SoloMid</i>	-1.333
<i>Top Esports</i>	0.2
<i>Unicorns Of Love.CIS</i>	0.214
<i>V3 Esports</i>	-1.8

Table 5: WCS Gold Advantages Statistics

	Average Gold Difference at 15mins
<i>DAMWON Gaming</i>	1939.647
<i>DRX</i>	1876.333
<i>FlyQuest</i>	-662
<i>Fnatic</i>	-701.455
<i>G2 Esports</i>	-93.643
<i>Gen.G</i>	896.556
<i>INTZ</i>	-1944.2
<i>JD Gaming</i>	555.8
<i>Legacy Esports</i>	-1163.25
<i>LGD Gaming</i>	690
<i>Machi Esports</i>	-2598.833
<i>MAD Lions</i>	-731.9
<i>PSG Talon</i>	-1045.364
<i>Rainbow7</i>	-14.429
<i>Rogue</i>	-1238
<i>Suning</i>	45.316
<i>SuperMassive Esports</i>	75.333
<i>Team Liquid</i>	1284.364
<i>Team SoloMid</i>	-909.333
<i>Top Esports</i>	955.067
<i>Unicorns Of Love.CIS</i>	-1255.929
<i>V3 Esports</i>	-1686.4

Table 6: WCS & LCK Win Rates Changes

	Win Rate Changes (%)
<i>DAMWON Gaming</i>	14.7
<i>Gen.G</i>	-14.2
<i>DRX</i>	-19.8

Table 7: WCS & LCK KPM Changes

	KPM Changes	Time to Gain Kills Changes (%)
<i>DAMWON Gaming</i>	0.11	-19.298
<i>DRX</i>	-0.053	14.846
<i>Gen.G</i>	-0.043	10.591

Table 8: Model Estimations and Errors

	Rank	Estimated Win Rates (%)	Actual Win Rates (%)	Errors (%)
<i>DAMWON Gaming</i>	1	77.9	82.4	-4.5
<i>Gen.G</i>	2	62	55.6	6.4
<i>DRX</i>	3	44.1	44.4	-0.3
<i>SeolHaeOne Prince</i>	4	15.2	0	15.2
<i>Afreeca Freecs</i>	5	12.5	0	12.5
<i>Griffin</i>	6	2.25	0	2.25
<i>KT Rolster</i>	7	0	0	0
<i>Hanwha Life Esports</i>	8	0	0	0
<i>SANDBOX Gaming</i>	9	0	0	0
<i>T1</i>	10	0	0	0
<i>Team Dynamics</i>	11	0	0	0
<i>Seorabeol Gaming</i>	12	0	0	0