Coin Toss: Your Odds of Winning

Joel Henriques

The University of Oklahoma

I. Abstract

The purpose of study is to explore whether winning the coin toss in sports can be utilized as an indicator in predicting the probability of winning the game. Unfortunately, the majority of sports do not typically retain records of the results of the coin toss due to the lack of significance (until this paper!). Thus, it is necessary to examine a sport that has robust data on either the coin toss or various characteristics that can enable the assumption of whether a player or team probably won the coin toss. The perfect sport happens to be tennis due to the quality of data recorded of many different aspects of practically every match in history. These aspects are the following: number of aces made by each player, the number of doubt faults, match duration, and etc... The aspect that is most significant to this study is the player who first served. I make the assumption that tennis players who win the coin toss in tennis will also serve first. This is a common etiquette in tennis and will be used as the main determinant for those who won the coin toss for each match. Ultimately, the research proposes that those who win the coin toss are substantially advantaged against those who did not.

II. Introduction

There has been much debate in sports research on whether the coin flip truly matters. Many sports such as college football, volleyball, tennis, and many other sports all use this coin toss method to determine which team or player shall be able to make the first decision of the match. For example, the player who wins the coin toss in tennis, also has the ability to choose to serve first or forego the option. Now, the first decision can be offset by the other party getting to choose the second decision of the match. It appears to be a method of equalizing the playing field. For instance, college football allows the team who won the coin flip to decide whether to kick – off or receive the ball first, while the other team gets to select which part of the field to reside. This is quite common in sports; however, many still debate the magnitude of the initial decision that is won by the coin flip. According to Victor Mather (2015), a renowned columnist for the New York Times, "In cricket … the toss can be vital. The pitch, the grassy area where the ball is bowled, can change as the game goes along." This comment from Mather illustrates that winning the coin flips matters tremendously in some sports due to the nature of the game such as the field conditions worsening as game progresses. Those who win the coin toss are able to reap the rewards of having an unscathed field in cricket, which can make all the differences.

In another article "Is There an Advantage To Serving First," Jeff Sackmann (2013) emphasizes that those who served first also won on average 52% of the time. Although this appears minuscule due to the fact that tennis is between two players who have a 50% chance of winning, Sackmann's research is suggesting that there are biases towards those who make the first decision of the match. Theoretically, if two identically talented players were playing one another, the differences in the outcomes could come from the advantage of winning the coin flip. The two identical players would be evenly matched, but one has the advantage of serving first which neglects the other player from having the same opportunity. Therefore, this study will continue to explore the advantages or disadvantage behind winning the coin flip in the sport of tennis through use of econometrics.

III. Description of Model

I have crafted a bivariate model that regresses wins on first serves (those who won the coin toss). The dummy variables are used to quantify whether those who served first, either won or lost. The wins would contain the value of one while loses would be held as zero. The *First.Serve* variable is also a dummy variable that quantifies which player served first or won the coin toss. There are 86,430 observations being used for this linear regression, which should be a sufficient amount of observations for seeing the overall impact that winning the coin toss can have on the likelihood of winning. I also included a multivariate model to see other interesting variables that might heavily contribute to winning such as aces and break points converted.

Model 1: ***(Units Wins)***

$$Wins = \beta_1 + \beta_2 First.Serve + \varepsilon$$

Model 2: ***(Units Wins)***

$$Wins = \beta_1 + \beta_2 First.Serve + \beta_3 Aces + \beta_4 Break_{Points_{Converted}} + \varepsilon$$

The chart listed below will describe each of the variables.

Figure 1 – *Defining Variables*

Wins	Dependent variables – Dummy Variable
First. Serve (Coin Toss Winner)	Dummy variables
Aces	Number of Aces made by each player
Break Points Converted	Tied breaks won by each player

The model should logically and naturally convey the effects of Y given X in terms of percentages since the variables being used are between 0 and 1. Thus, there is no need to use Log in order to convey the data into percentages.

To reiterate, this model does assume that individuals who served first were also the ones who benefitted by winning the coin toss. My hypothesis is that players who win the coin toss are more likely to win the match, since they have the advantage of choosing the first decision in a match or game. In order to test this hypothesis, a positive correlation between the variable *First.Serve* and *Wins* would be sufficient to conclude that winning the coin flip is crucial to the outcome of these tennis matches. These models will aid comprehension and compute the degrees to which winning the coin toss can positively or negatively impact the player's results.

A Logit or Probit model could have been a more efficient method of observing the data rather than utilizing OLS, but the measurement of interest would have been the average marginal effect produced by those methods. Although extracting the average marginal effect from these models in order to understand the wins (dummy variables) used as the dependent variable in this model is effective, the values of the average marginal effect would have been extremely similar to the intercept of the OLS regression. Thus, I did not see the significance of running a Probit or Logit regression for this model.

Model 2 encompasses a few more interesting independent variables to control for such as number of aces by each player, break points converted, and first serves. The notion behind controlling for more variables is to observe the results of the coin flip compared to other aspects of the matches.

IV. Description of the Data

The data was collected from a website titled "ATP World Tour Tennis Data." For the purposes of this study, it seemed vital to accumulate thousands of observations on first serves and wins. Otherwise, a smaller data set can exacerbate or skew the data significantly. The data contains tennis matches between the years of 1991 – 2016 and is cross – sectional data. The data was also combined by the winner and loser statistics per match recorded and did not have available information on those who served first.

Therefore, I was able to manipulate the data by creating a new column that records the differences in the winner's total games served and returned. A negative number would indicate that the winner did not serve, a zero meant that it was inconclusive on which player served, and positive numbers indicated that the winner served first. Then, I removed the inconclusive serving data (all the zeros). The column for serving data of winners only contained negative or positive whole numbers, but I changed all of the negatives to zeros. This is to distinguish the differences between the winners who won the coin toss from the winners who did not. The remaining positive values indicated winners who served, and then I created a new column for losers who served. From this, I was able to split the winner's data from the loser's data and move the loser statistics underneath of the winner's statistics. Now, I had a column for those who served or did not serve and built a binary (1 or 0) column for those players that either won or lost. This enabled me to run a linear regression on the dummy independent variable (First.Serve) on the dummy dependent variable (wins). All regressions calculated will be reported with robust standard errors in order to control for the distribution of the error term, since it is unlikely that u is equal to zero.

V. Interpretation of the Variable and Estimated Model

The interpretations for Figure 2 & 3will be in terms of percentage change of prices in the presence of the changes in the explanatory variables ($\Delta y \approx (100\beta_1)\Delta x$). These are the results for both models:

Figure 2 – *Regression Results for Model 1 (Unit Wins)*

```
> est.rob <-lm_robust(Wins ~ First.serve, data = df)
> tidy(est)
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)    0.316   0.00224      141.       0
2 First.serve    0.367   0.00316      116.       0
> glance(est)
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df  logLik     AIC     BIC deviance df.residual
      <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>   <dbl>   <dbl>   <dbl>    <dbl>       <int>
1     0.135         0.135 0.465    13486.       0     2 -56464. 112934. 112962.   18691.       86428
```

Figure 3 – *Regression Results for Model 2 (Unit Wins)*

```
> est.rob <-lm_robust(Wins ~ First.serve + aces + break_points_converted, data = df)
> tidy(est)
# A tibble: 4 x 5
  term                   estimate std.error statistic   p.value
  <chr>                     <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)             -0.0804  0.00266      -30.2 5.04e-199
2 First.serve              0.268   0.00258      104.  0.
3 aces                     0.0113  0.000267      42.3 0.
4 break_points_converted   0.137   0.000644     213.  0.
> glance(est)
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df  logLik    AIC    BIC deviance df.residual
      <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>   <dbl>  <dbl>  <dbl>    <dbl>       <int>
1     0.446         0.446 0.372    23221.       0     4 -37184. 74378. 74425.   11964.       86426
```

The difference in R^2 in Figure 3 opposed to Figure 2 is due to the fallacies associated with the nature of using R^2. As more independent variables are introduced into the model, R^2 will continue to increase regardless of the relevancy of each variable. Figure 2 has an R^2 of 0.135, which means that the first serve explains roughly 13.5 percent of the wins (the dependent variable). While Figure 3 reports an R^2 of 0.446. This means that the independent variables of number of aces each player had, number of break points won, and serving first can explain 44.6% of the wins (dependent variable).

In Figure 2 the intercept calculation is suggesting that each player has a 31.6% chance of winning the game when they first step on the court, without taking into any considerations of the changes that the independent has on the dependent variable. For every unit increase in First serves (or wins of coin tosses), will increasing the winning probability by 36.7%. The *First.serve* variable is statistically significant at the 5% level.

To dive deeper into other interesting analysis, Figure 3 has an intercept coefficient of -0.08. Meaning, that each player has a negative 8% chance of winning once entering the court. If the player wins the coin flip, then their probability of winning the game increases by 26.8%. For every additional ace a player obtains, increases their probability of winning by 1.1%. Although aces appear to have miniscule effects on the outcomes, there can be up to 30 plus aces made by a player in a game. An additional break point won by a player, will increase their odds of winning by roughly 13.7%. The independent variables aces obtained, break points converted, and first serves were all statistically significant at the 5% level.

VI. Economic Analysis and Evaluations

The linear regression from Model 1 is suggesting that all individuals have practically 1/3 probability of winning, without considering any of the conditions of the independent variables. This exacerbates the effects that winning the coin toss has on the outcomes of the match. For example, the player who wins the coin toss increases their probability of winning to roughly around 67% ((0.316) + (0.367*(1))). Therefore, model 1 puts tremendous emphasizes on the necessity to win the coin toss.

Model 2 accounts for more of the characteristics that the players can control such as their ability to obtain aces and win break points. The most interesting aspect of model 2 is the intercept coefficient being -8%. The effects of winning the coin toss has a less impactful effect on outcomes which decreased to 26%, however, this is still quite substantial. To further help conceptualize the significance of winning the coin toss let's assume the close to worst-case scenario where a player did not win the coin toss, but at the very least was able to score one ace and convert one break point opportunity. This individual would only have a 6.8% (-8%+((0)*(26.8%))+((1)*(1.1%))+((1)*(13.7%)) chance of winning the game. Whereas, a player who is also having a terrible match but wins the coin toss would has an 33.6% (-8%+((1)*(26.8%))+((1)*(1.1%))+((1)*(13.7%)) chance of winning. This is a percentage change of 394% ((33.6-6.8)/6.8) for the players who won the coin toss as opposed to those who did not.

VII. Summary

In conclusion, there appears to be strong evidence that supports the notion that parties that win the coin toss will have a substantial advantage over those who did not. The advantages will be most apparent in the circumstance where competitors have equal ability. However, many individuals will argue that over the aggregate of a season, both parties have the same opportunities to receive the benefits to the point that the rewards will not exceed the loses. This originates from the theory of "regressing to the mean." Contrary to popular belief, the issues truly arise in sports such as tennis, where there is no real season games and only high-pressure moments such as placing in tournaments. For a sport such as tennis where the coin toss can almost determine 1/3 of the probability of winning, the player who loses the coin toss is at a substantial disadvantage compared to their opponent. Ultimately, if the effects of the coin toss are the same across all sports, then the coin toss can truly make the difference between a National Championship and a loser.

VIII. Works Cited

Datopian. (n.d.). ATP World Tour tennis data. Retrieved November 15, 2019, from https://datahub.io/sports-data/atp-world-tour-tennis-data#resource-match_stats_1991-2016_unindexed.

Mather, V. (2015, November 30). Coin Toss Retains Its Place in History, if Not in Cricket. Retrieved December 2, 2019, from https://www.nytimes.com/2015/12/01/sports/international/coin-toss-is-secure-in-sports-lore-if-not-in-cricket.html.

Sackmann, J. (2013, October 25). Is There an Advantage To Serving First? Retrieved November 15, 2019, from https://www.tennisabstract.com/blog/2013/10/24/is-there-an-advantage-to-serving-first/.