Football Championship Ticket Options

Introduction

Since its adoption in the 2014-2015 season, the College Football Playoff (CFP) has given college football fans even more reason to be excited about rankings. No longer did a team need to be in the top 2 of the Bowl Championship Series (BCS) to be a championship contender – instead, the top 4 teams entered into a playoff. In capitalizing on the intersection between excitement and fandom, the CFP teamed with OptionIt to introduce an RSVP system for championship game tickets, starting in 2018. A fan may by an RSVP (an option) for a team. If that team would make it to the championship game, then the fan holding the option would be able to purchase a ticket at face value. In the end, this is likely to save fans a considerable amount of money as compared to purchasing tickets on the secondary market – tickets on the secondary market were estimated to be an average of $2,689 for the 2018 Championship game. In financial modeling terms, this system could be considered a call option for speculation. When buying an option (the RSVP), the buyer has the right to buy a stock (a ticket) in the future at the strike price (the face value of the ticket).

There is certainly a huge amount of speculation built into this – at present an option for a lower-level seat for Alabama is very nearly $600. This is a considerable financial risk when purchasing such an option. Coupled with a possible face value around $800, a fan buying an option for $600 could easily have $1,400 wrapped up into a ticket. This is still less money than what would be spent on the secondary market, so the reward is potentially high.

A source of inquiry is what drives the price of these RSVPs. Before the season begins, would knowing a teams performance during the previous year drive prices? Would preseason ranking have any influence on prices? After the start of the season, would rankings, performance, injuries, or other variables affect prices? These questions will be explored throughout our paper.

General Research Questions

When examining the option prices, it becomes apparent that the best teams have the highest prices; conversely teams with little chance of making the CFP have very low dollar values. Among those top teams, though, what might be driving the prices? As a preliminary pass, we are going to look at current values and volatility, and what impact wins, injuries, strength of schedule, and rankings have on those metrics. Specifically, we would anticipate that volatility has a positive relationship with wins (i.e., the more wins a team has, the more the option changes), but a negative relationship with rankings (the lower the ranking, the more the option changes). Number of injuries, while certainly challenging to a given team, might not be critical at this point within the season. Nonetheless, we would anticipate that the number of injuries will drive the option value down. We can certainly imagine a world in which strength of schedule has an impact on prices and volatility – a nightmarish schedule might all but destroy your chances of making it to the playoffs.

Data

The internet is full of interesting data and finding it is never too hard. While finding interesting data does not present a challenge, finding data that is free to use can be tricky. In the context of data, free to use does not necessarily mean free in a monetary sense. Instead of a paywall, data is often locked behind terms of service (ToS). While scraping data has become easier, data creators are getting more sophisticated at "hiding" the data (i.e., it is not sitting nicely within the html, but loads through JavaScript) or ToS. With the appropriate acumen, the location of the data ceases to pose a challenge; the same cannot be said for ToS. In scraping our data, we not only read the ToS, but we also examined each site's robots.txt. Since we wanted to conduct our data collection in a programmatic fashion, we wanted to ensure that we were not going to be running afoul of restrictions placed within the robots.txt file. To that end, all the data that we collected is free to use – in other words, no restrictions apply.

Sources

Given the research questions, we sought to collect data from several different sources. Given the excellent compilation of team rankings, Ken Massey's website was used to amass rankings.

Injury counts were pulled from Don Best's website and projected winners were pulled from Odds Shark.

Finally, the RSVP data was pulled from the json object on the CFP-RSVP site. While the main page displays the current option price for an upper-level ticket (UL_price), the json object also contains prices for lower-level tickets (LL_price). There are also different hotel packages within the data, but our primary focus will be on the two ticket prices. For visualizations, both LL_price and UL_price will be presented; for analyses, we will focus our attention for LL_price.

Issues and Limitations

The RSVP data presented a few challenges along the way – mainly that the data source changed at some point during the course of data collection. This data source change resulted in a considerable span of data loss. Further, issues arose due to the automatic nature of data collection. Windows Task Scheduler was used to automatically run our data collection scripts daily – unfortunately, this will not run if a computer is shut down. Finally, new ratings were added to the Massey Ratings page. Given the position of the Massey ratings in our data join, many columns became displaced from their original position. While this was fixed, some values were not recovered.

Results

Previous Year-end Rank and Early Prices

As a starting point, we began by looking at Massey's year-end rankings and early option prices (February 2018).

In an effort to reduce visual clutter and focus on potentially high-value teams, we can visually inspect the top 25 teams at the end of the 2017-2018 season.
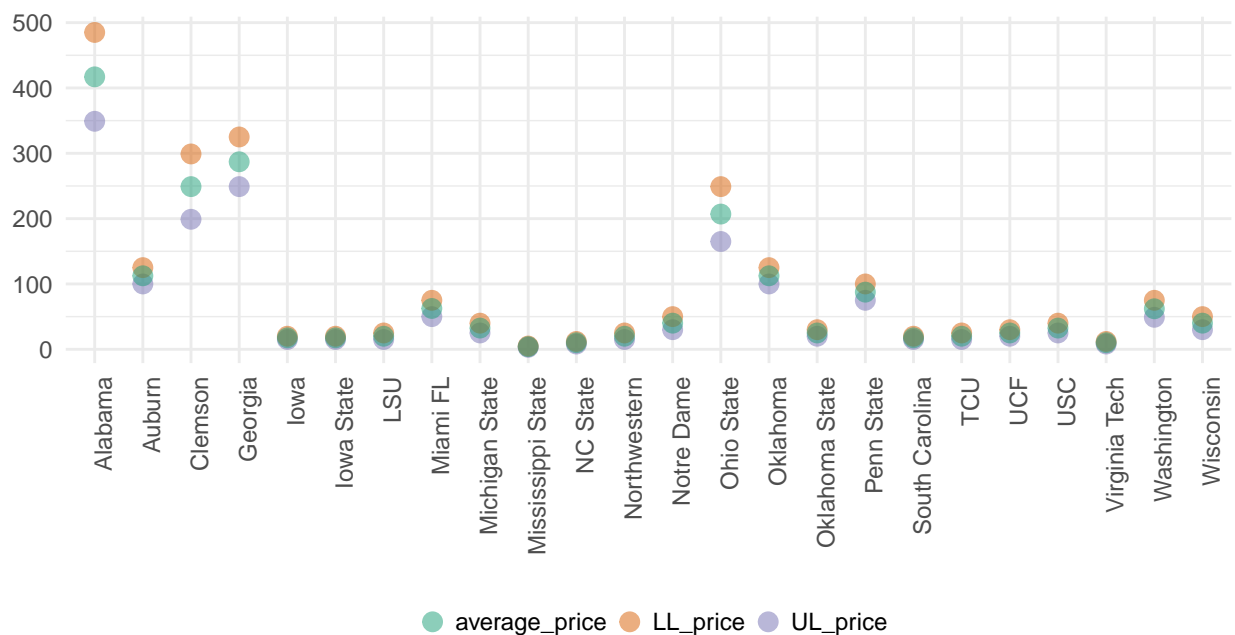

The preceding visualization provides a clear picture of how early prices reflect the previous year rankings.

Some teams, like Penn State, Oklahoma, and Ohio State, seem to have started out a bit higher than what the end-of-season ranking would have suggested. All of this is not terribly surprising, given that the previous year is all of the information informing pricing.

Ranking Correlations

For the sake of exploration and to limit the chances of future multicollinearity, we wanted to explore how some of the more well-known ranking were correlated. We would anticipate that they are generally well correlated, but :
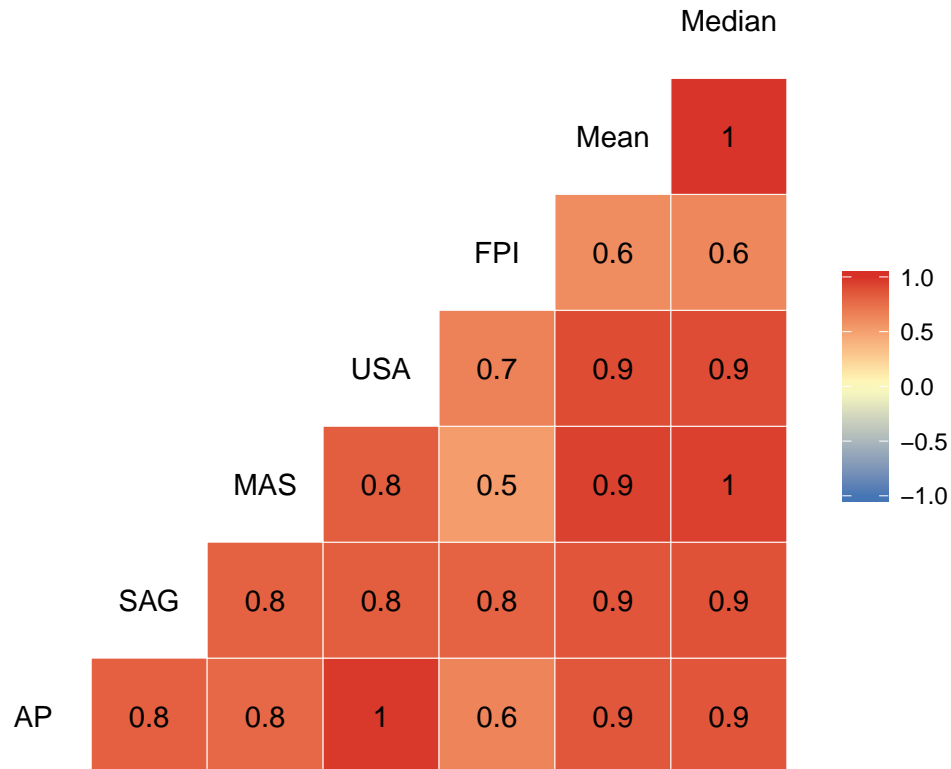
Generally, strong correlations are present. For upcoming models, we will opt to use the mean ranking. There are a few reasons for this decision: 1) rankings like AP and USA only go up to 25, 2) there are sometimes significant differences between the more common ratings (AP & USA Today Coaches) and analytic ratings like Massey and Sagarin, and 3) we want something that will capture all of the ratings present.

Top 25 Trends

The following figure shows the price trend for every team that has been in the AP top 10 since the beginning of the season.

A quick visual inspection of every team that has been in the top 10 this year might suggest persistent high confidence for Alabama, Clemson, and Georgia (but to a slightly less degree). Not surprisingly, Ohio State and Oklahoma are starting to trend upwards; meanwhile, Notre Dame has consistently high prices (relative to their real chances of actually making it to the playoffs), but on a potential downward trend.

Volatility

While the intervals in our option price series are not always equivalent, we do have many time points for each team. When dealing with any investment-related data, volatility is usually of interest. We can use standard volatility calculations to produce a volatility measures for each team's upper and lower ticket prices.
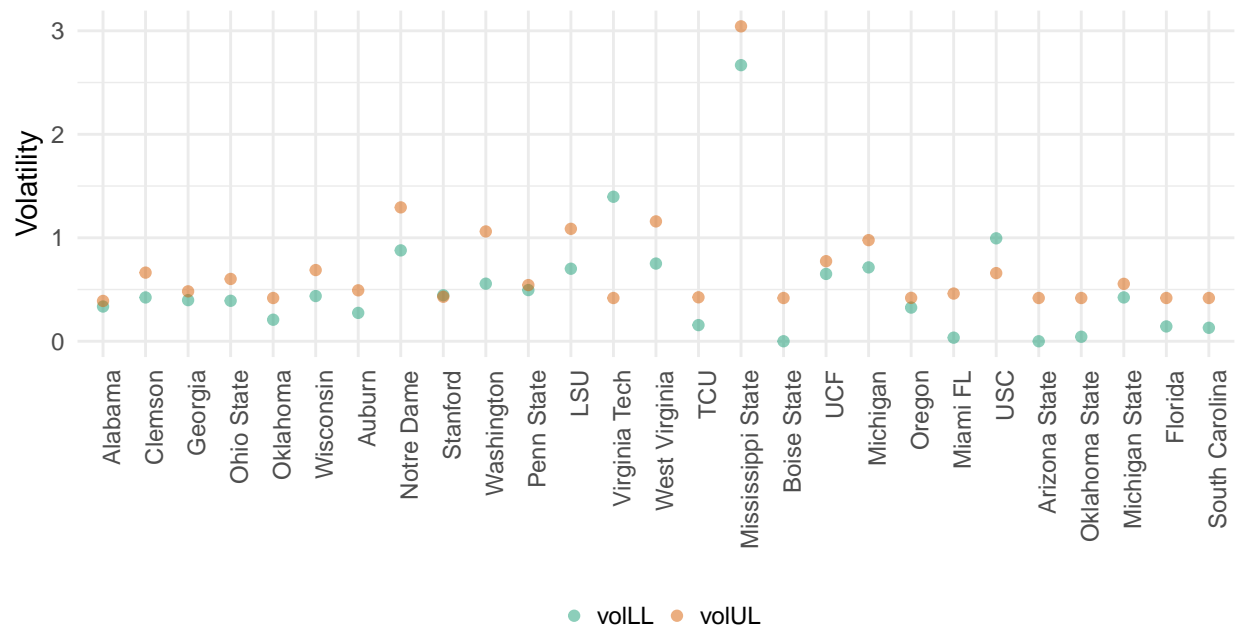
Mississippi State stands out in terms of volatility. Despite finishing around 14 in the 2017-2018 season, their preseason rankings were a bit below that. However, they are now beginning to rise in the rankings.

We can compare these plotted volatility values with the values that CFP-RSVP uses to graphically illustrate a team's price trend.

First, we can see which 25 teams have had the most price changes throughout the CFP-RSVP's existence. Note that the following visualizations displays the total number of price changes (includes both upper and lower level changes).

With some exceptions (Texas, Florida State, Arizona, Memphis, Miami, Northwestern, and N.C. State), we see that those teams with the most number of changes are well-aligned with the volatility within the teams that have been in the top 25 throughout the season.
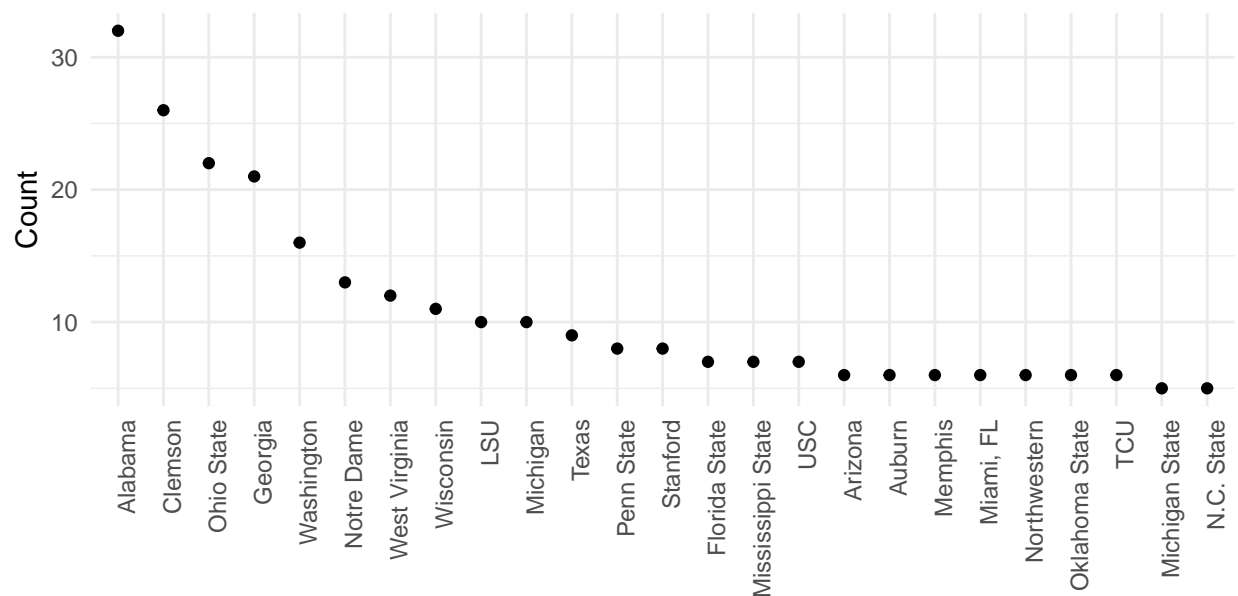
We can also plot the reported series for all of those teams which have appeared in the top 25 this season:

— lowerLevelPrice — upperLevelPrice
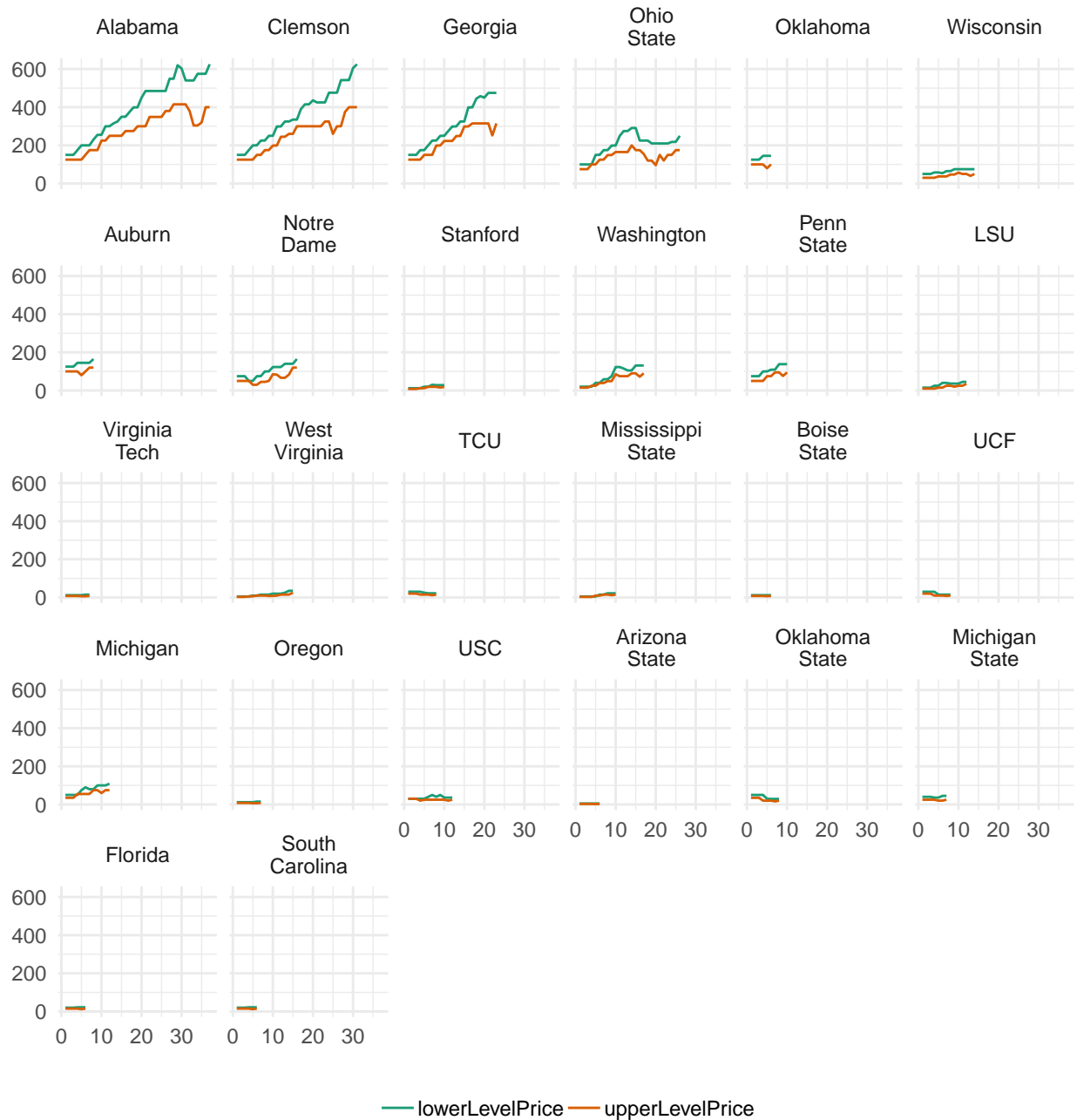
Unfortunately, there is no time component to this reported series – only a series of numbers. So finding out what caused Alabama's dip in the reported numbers would not be possible. Comparing this abbreviated series (essentially a step-like series) to our series with daily recordings is very interesting.

Explaining Volatility

We want to explore if current mean ranking (Mean), wins, injury count (n), and strength of schedule (masseySoS) could be predictive of volatility. Since we want to allow for as much flexibility in our model as possible, we will use generalized additive models for our models. While we will not smooth wins (there really are not enough values at present to effectively smooth it), all other predictors will be smoothed. For the sake

of space, we will just look at lower level (we do, however, expect that upper and lower levels will behave similarly).

term	edf	ref.df	statistic	p.value
s(Mean)	1.728111	2.152087	3.945309	0.0220661
s(n)	8.446084	8.893720	2.920634	0.0029248
s(masseySoS)	1.000000	1.000000	0.027963	0.8676240

term	estimate	std.error	statistic	p.value
(Intercept)	0.3913439	0.1502452	2.604701	0.0109873
wins	-0.1193643	0.1122825	-1.063071	0.2909956

For our smoothed terms, we can see that strength of schedule does not do much, but mean ranking and number of injuries do contribute to volatility. Mean ranking behaves as anticipated (a near linear negative relationship), but the number of injuries has an interesting relationship with volatility (see the following plots). Wins, our lone parametric term, did not have much to contribute to our model.
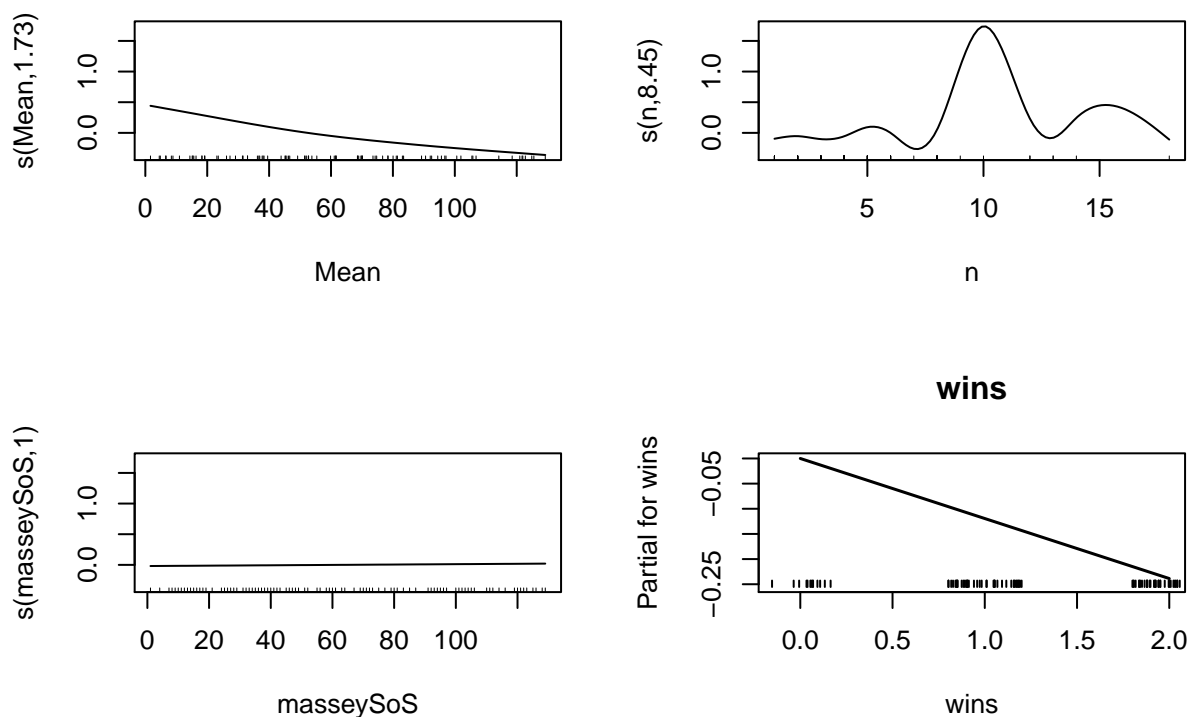
### Explaining Current Prices

We will run a similar model, except using current lower level price as our outcome variable.

term	edf	ref.df	statistic	p.value
s(Mean)	8.349139	8.873551	22.006746	0.0000000

term	edf	ref.df	statistic	p.value
s(n)	3.871142	4.705951	2.686935	0.0269568
s(masseySoS)	1.000000	1.000000	0.711355	0.4016005

term	estimate	std.error	statistic	p.value
(Intercept)	49.097829	17.01845	2.8849766	0.0050785
wins	-9.624918	12.67738	-0.7592199	0.4500457

While wins and strength of schedule did not offer much (again), mean ranking and number of injuries did offer some predictive power. The effect of mean ranking on the current price took on an interesting pattern (as seen below), while the number of injuries had a negative relationship (as would be expected).
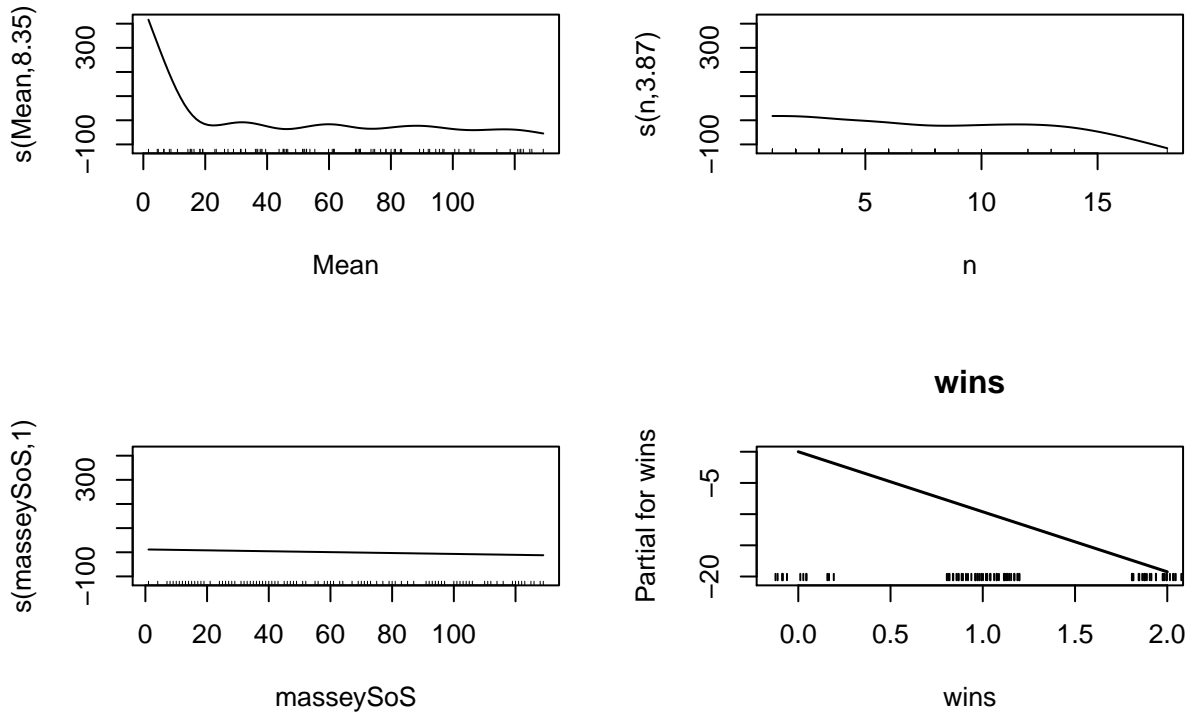
Model Testing

In an effort to see how our model predicting lower level volatility might perform in a training/testing scenario, we split our data 50%/50%.

The following are error statistics from our trained model predicting new data:

RMSE	0.4743836
Rsquared	0.2612521
MAE	0.2509689

Perhaps not the greatest performance between training and testing, but certainly nothing to completely disregard (reasonable R^2 given our context and a mean absolute error that isn't embarrassing).

In doing the same 50%/50% split for current price, we have the following results:

RMSE	119.1547475
Rsquared	0.0015035
MAE	51.4062730

These results are not as promising as those generated with our volatility models.

Take Away Messages

The models presented here are incomplete, given that we are only 3 weeks into the college football season. Despite this, we are encouraged that the ticket options are behaving in a predictable manner. As data continues to roll in, our models will grow in complexity and predictive power. Ultimately, the goal is buy before a team becomes the hot dark horse and sell become a team's dreams burn in flames.

While we don't have data back to the inception of CFP-RSVP, we do have a pretty broad swath. Given our visualizations and our models, the strategy is clear: buy early and buy towards the top of the projected rankings. There is no such thing as parity within college football; of the possible 16 slots combined over the last 4 years, there have been 10 unique teams in the playoffs (with Alabama making 4 appearances, Clemson making 3 appearances, and Ohio State and Oklahoma each having 2 appearances). We would likely invest early in the teams that are likely to be around when the playoffs start – even the casual fan can have a pretty good shot at picking the top.