

Potential Week 4 NFL Model

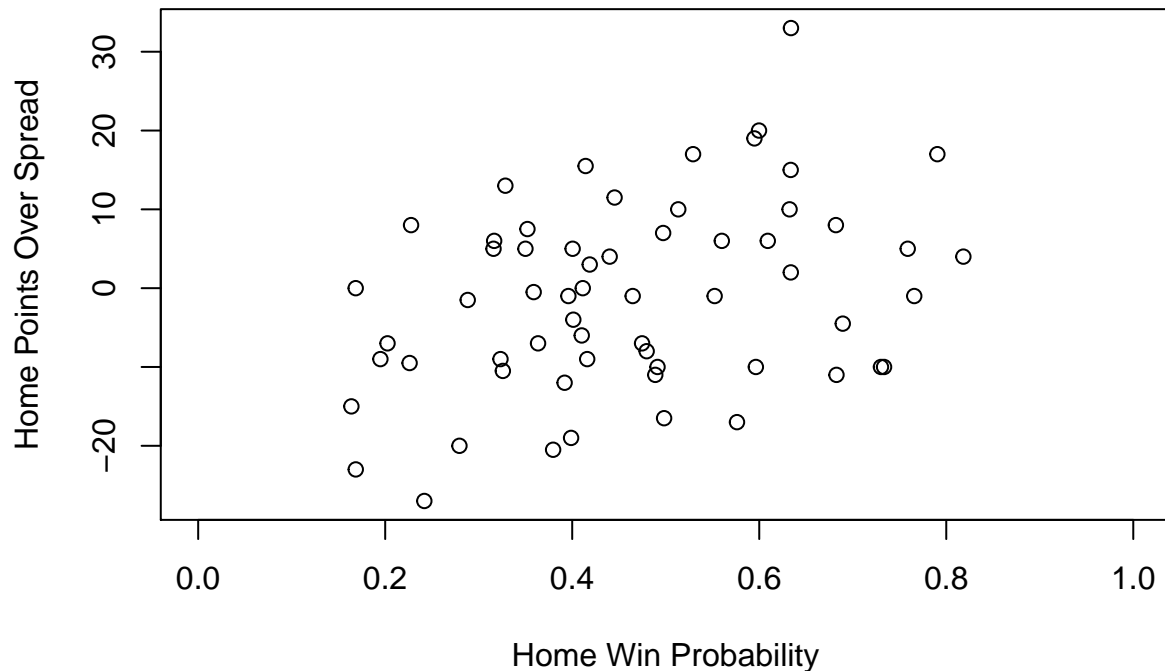
James Whedee

October 14, 2015

Exploratory Analysis

The following is exploratory analysis of a Logistic regression model created after Week 4 of the 2015 NFL season. It is based off of the most recent weeks Offensive and Special Teams DVOA.

The following is a plot of the number points by which the home team beat the spread against the model's predicted probability of the home team winning the game.



The linear relationship between the two is quite strong. The interpretation of the coefficient is that for every .01 increase in the model's predicted probability of the home team winning, the expected margin by which the home team beats the spread increases by .24 points.

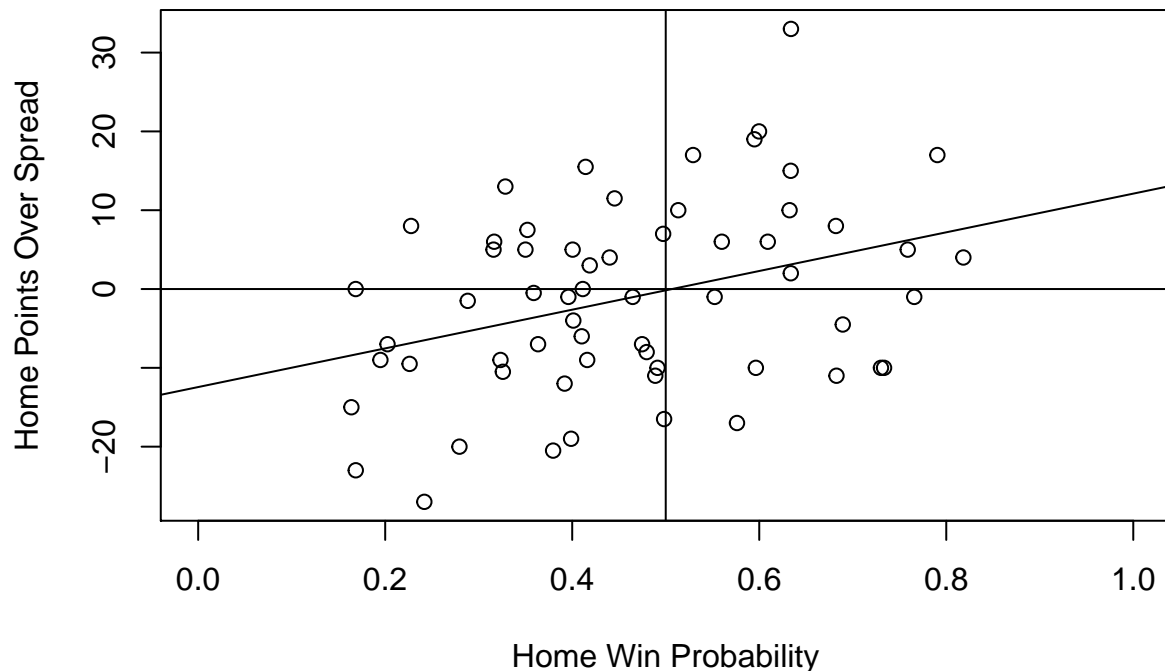
```
##
## Call:
## lm(formula = VOffVSpec$Home.Cushion ~ VOffVSpec$P.Home)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.496  -7.353  -1.114   8.305  29.882
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -12.440     4.198  -2.963  0.00439 **
## VOffVSpec$P.Home  24.543     8.519   2.881  0.00552 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 11.27 on 59 degrees of freedom
## Multiple R-squared:  0.1233, Adjusted R-squared:  0.1085
## F-statistic: 8.3 on 1 and 59 DF,  p-value: 0.00552
```

This model has several promising characteristics. For one, it has a very reasonable distribution of predicted winning probabilities, with most games a bunched in the middle 50%.

```
##          0%          25%          50%          75%          100%
## 0.1638350 0.3499787 0.4400557 0.5963987 0.8182847
```

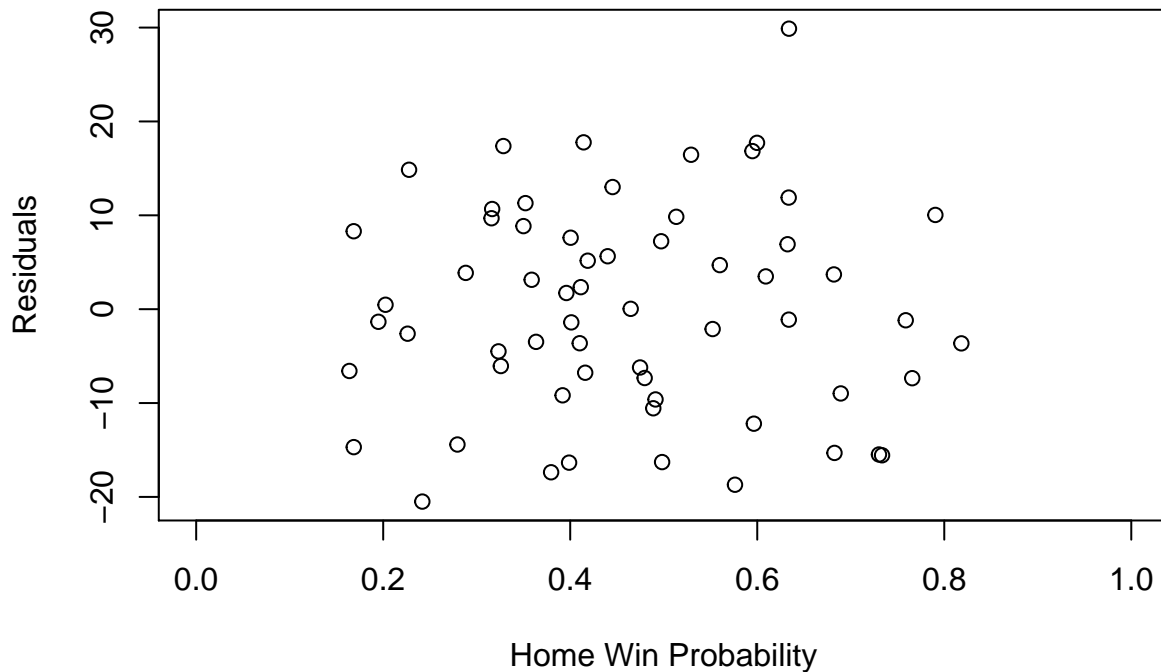
Also promising for this model is how closely it fits to the point (.5,0) which is intuitively interpreted as a 50/50 prediction indicating that the teams would push.



Finally, this model has residuals which diminish over time, which would be expected from a predictive model as DVOA values become less variable.

```
##
## Welch Two Sample t-test
##
## data:  earlyWeeks and laterWeeks
## t = 1.9914, df = 57.68, p-value = 0.05118
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.01629977  6.16779103
## sample estimates:
## mean of x mean of y
## 10.659663  7.583917
```

Here is a plot of the residuals showing no discernible underlying pattern.



Betting Strategy

We can get an idea of “safe” bets by taking into account the current expectations for margins of error based on our residuals. Shown below are quantiles for the absolute values of the residuals in the “later weeks” where the residuals have decreased.

##	40%	50%	60%	70%	80%
##	5.372758	6.770534	8.741358	9.899144	12.522833

Absolute value was taken to be conservative, since we do not know which way the residuals will differ (in or against our favor). Since the residuals appear to be normally distributed, one could devise a more aggressive (and likely more accurate) system around the idea that only half of the expected residuals would cause a loss.

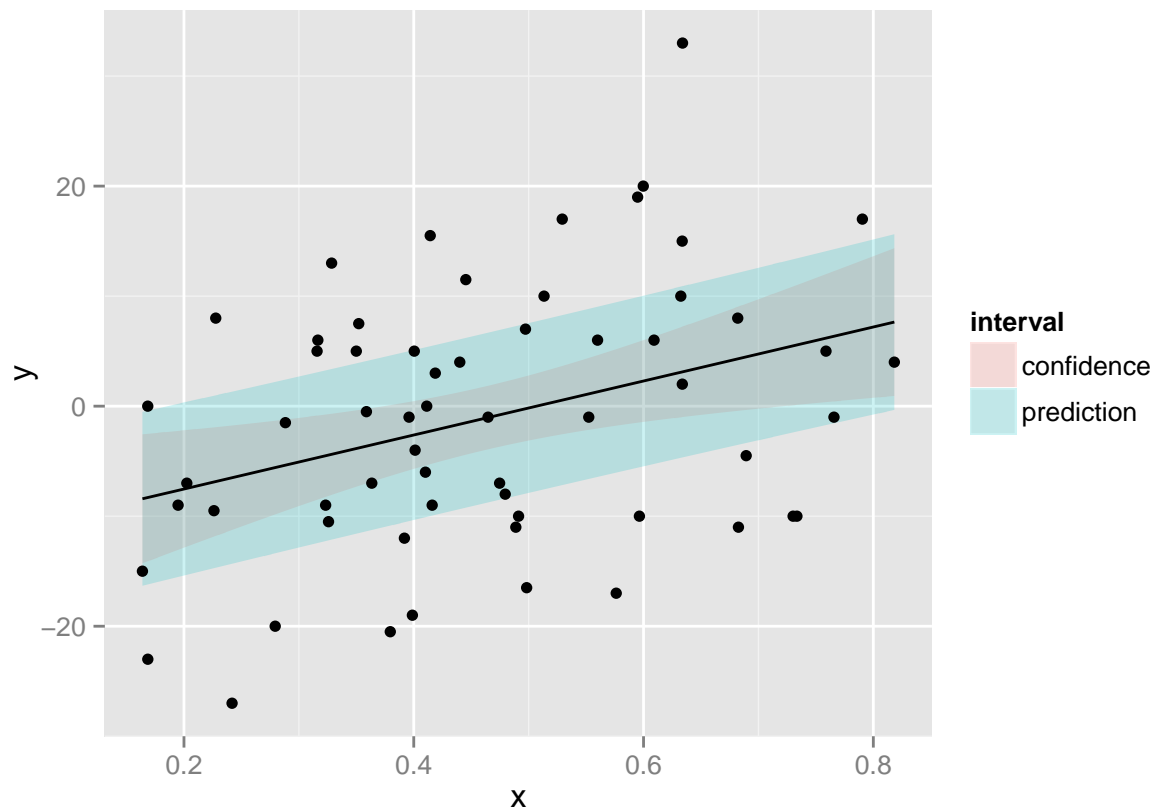
In other words, these quantiles tell us that allowing ourselves an 8.74 point cushion will result in roughly a worst case scenario of a loss of 40% of our bets and in roughly an average case scenario of a loss of 20% of our bets.

So, if we are willing to accept a 50% loss in the worst case with an average case of 25% losses, we would be looking for predicted winning probabilities that map to ± 6.77 .

These would be home team winning probabilities of less than 23.1% and more than 78.3%.

In this data set, that occurred in 9 of the 61 games and resulted in a betting record of 7-1-1.

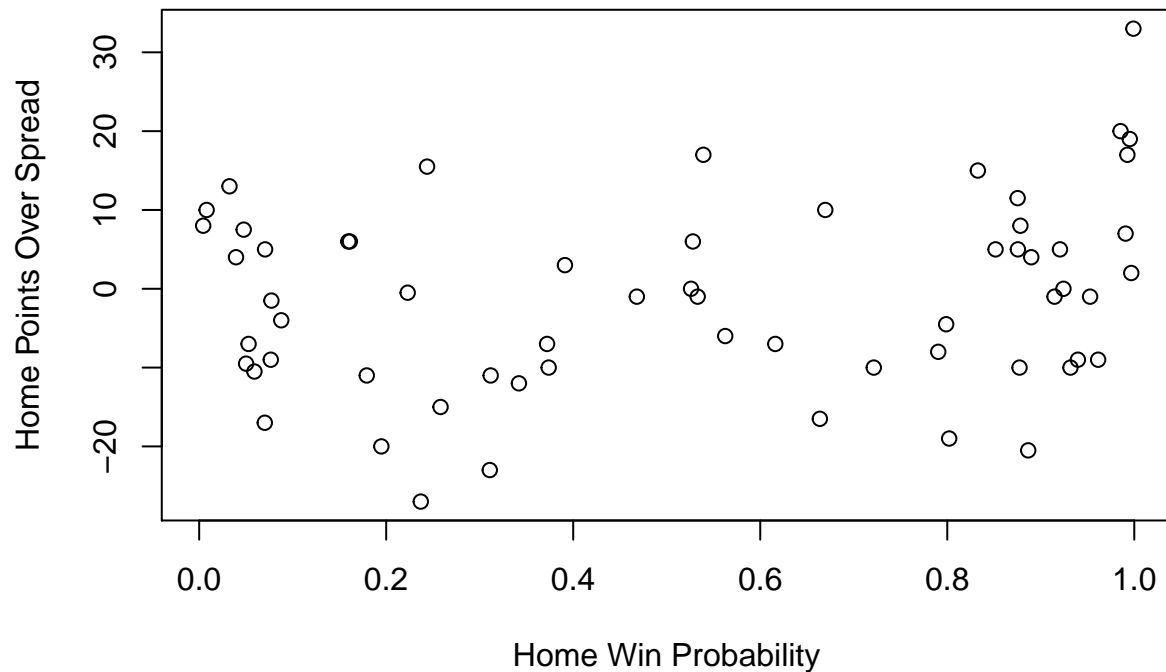
Further investigation would be needed to find the optimal cut-offs. Here’s a nice visualization of prediction intervals at .5 (which would yield 25% losses). This indicates we will have 25% losses all games with predicted winning probabilities less than 18.3%



```
##          lwr          upr          x
## 110 -16.32397 -0.51376630 0.1638350
## 210 -16.15346 -0.35979397 0.1704456
## 310 -15.98313 -0.20564513 0.1770562
## 410 -15.81297 -0.05131923 0.1836668
```

Comparison to Equally Performing Model

Finally, for comparison here is a similar attempt at this analysis on a logistic regression which is performing equally well to show that the behavior exhibited by this model is not simply a consequence of its current performance.

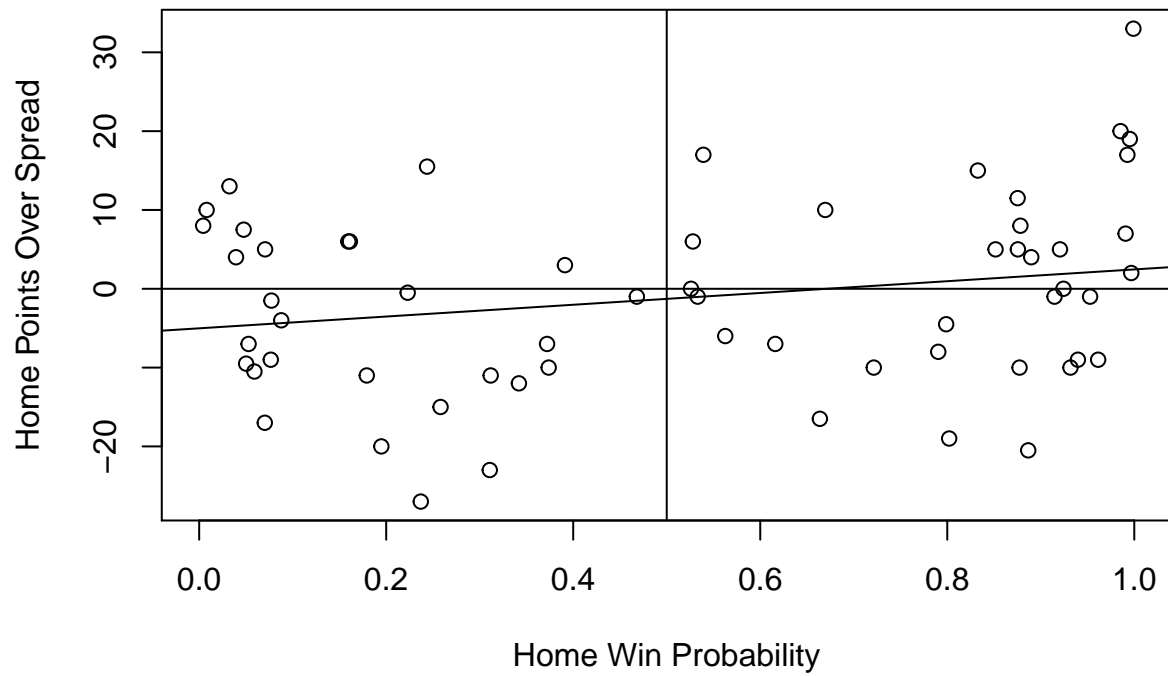


This model has a very unreasonable distribution of predicted winning probabilities, with most games a bunched at the extremes.

```
##           0%           25%           50%           75%           100%
## 0.004286402 0.179211469 0.533076871 0.878229840 0.999060071
```

This model has a less convincing line of best fit, which is nowhere near the pivotal (.5,0) point.

```
##
## Call:
## lm(formula = H0ffVSpec$Home.Cushion ~ H0ffVSpec$P.Home)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.7534  -8.3120   0.0338   9.4909  30.5513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.018     2.687  -1.867   0.0669 .
## H0ffVSpec$P.Home  7.473     4.231   1.766   0.0825 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.73 on 59 degrees of freedom
## Multiple R-squared:  0.05023,    Adjusted R-squared:  0.03413
## F-statistic:  3.12 on 1 and 59 DF,  p-value: 0.08251
```



Here is a plot of the residuals showing how useless the regression is. There is almost no difference in the residuals and the original plot due to the weak slope of the line.

