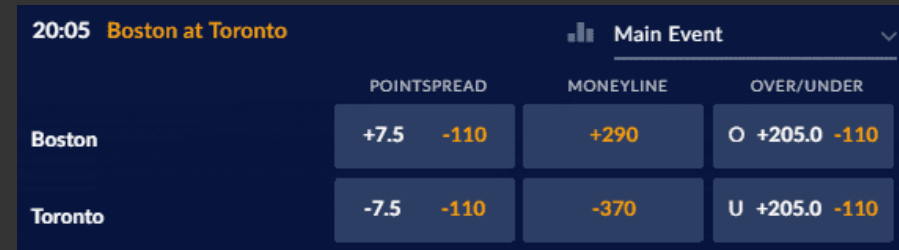


**LIFE'S TOO SHORT TO BET THE
UNDER:**
PREDICTING OVERS IN THE NBA

Spencer Kerch

OVERVIEW

- A very common bet in sports gambling is the total
 - If total points scored is over/under a certain threshold
- Sportsbooks want to set the total as close to what will happen as possible
 - This is to get equal money on the over as the under
- Gambling is meant to be difficult
 - Even being just above 50% is a success
- The Process:
 - Data Set
 - Training & Testing
 - Linear vs Non-linear



A screenshot of a sports betting interface. At the top, it says '20:05 Boston at Toronto' and 'Main Event'. Below this is a table with three columns: 'POINTSPREAD', 'MONEYLINE', and 'OVER/UNDER'. The rows are for 'Boston' and 'Toronto'.

	POINTSPREAD	MONEYLINE	OVER/UNDER
Boston	+7.5 -110	+290	O +205.0 -110
Toronto	-7.5 -110	-370	U +205.0 -110

<https://news.sportsinteraction.com/wp-content/uploads/2018/04/nba-over-under-example.png>

DATA SET

- Data from 2012-2019
 - Separated by season
 - Odds and game data
- Advantages
 - Can group by season
 - Able to get season averages
- Disadvantages
 - Basic stats
 - PTS, AST, REB, TOV, 3%, FT%, FG%, Wins, Total, Spread
 - Multiple sportsbooks



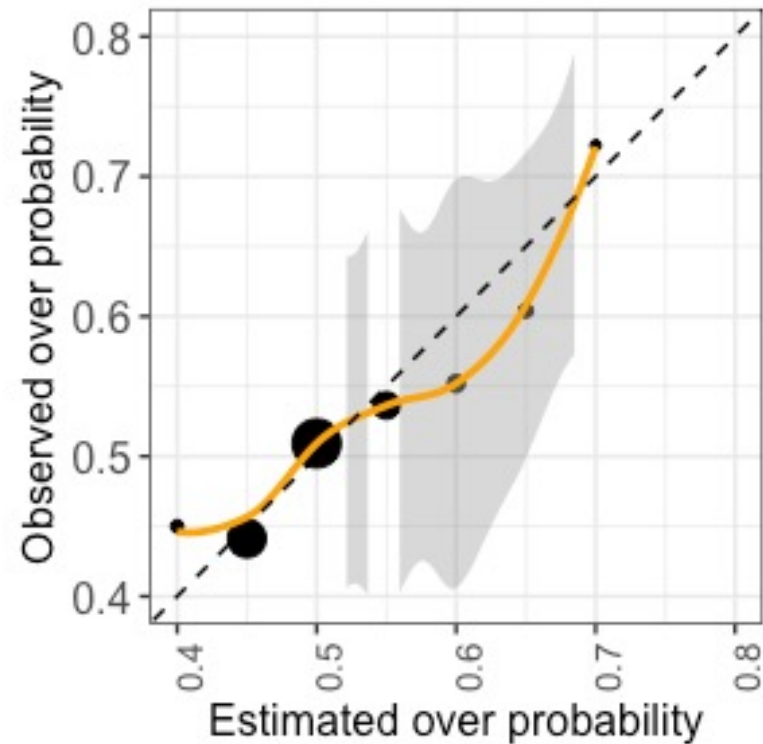
TRAINING: LOGISTIC REGRESSION MODEL

Logistic Model Performance

on the training set

Models	Features	AIC	AIC Difference	AIC Weight
Mod6	6	8,796.03	0.00	59.02%
Mod4	8	8,798.68	2.65	15.69%
Mod5	7	8,798.84	2.80	14.55%
Mod3	9	8,800.33	4.30	6.88%
Mod2	10	8,802.15	6.11	2.78%
Mod1	11	8,804.02	7.98	1.09%

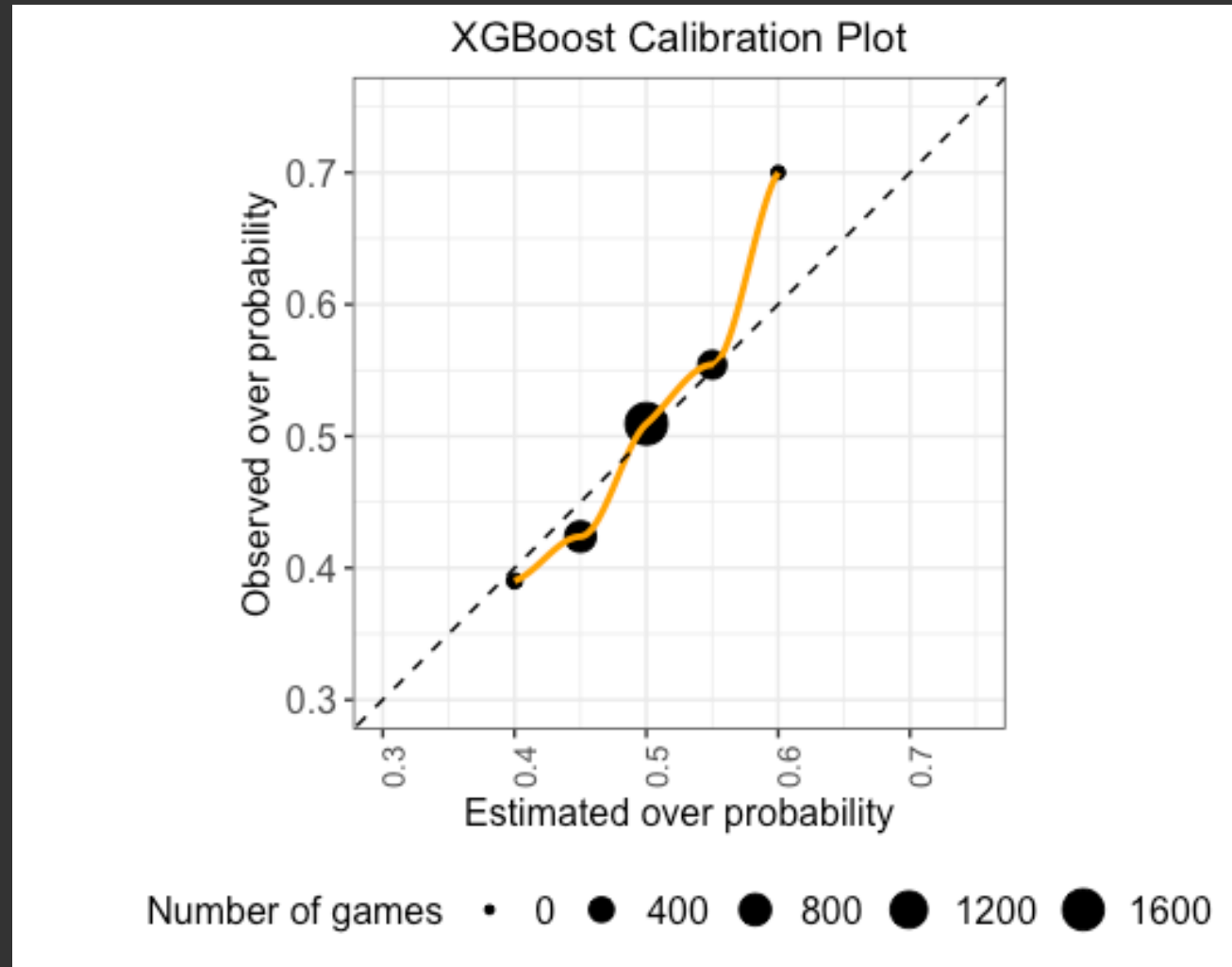
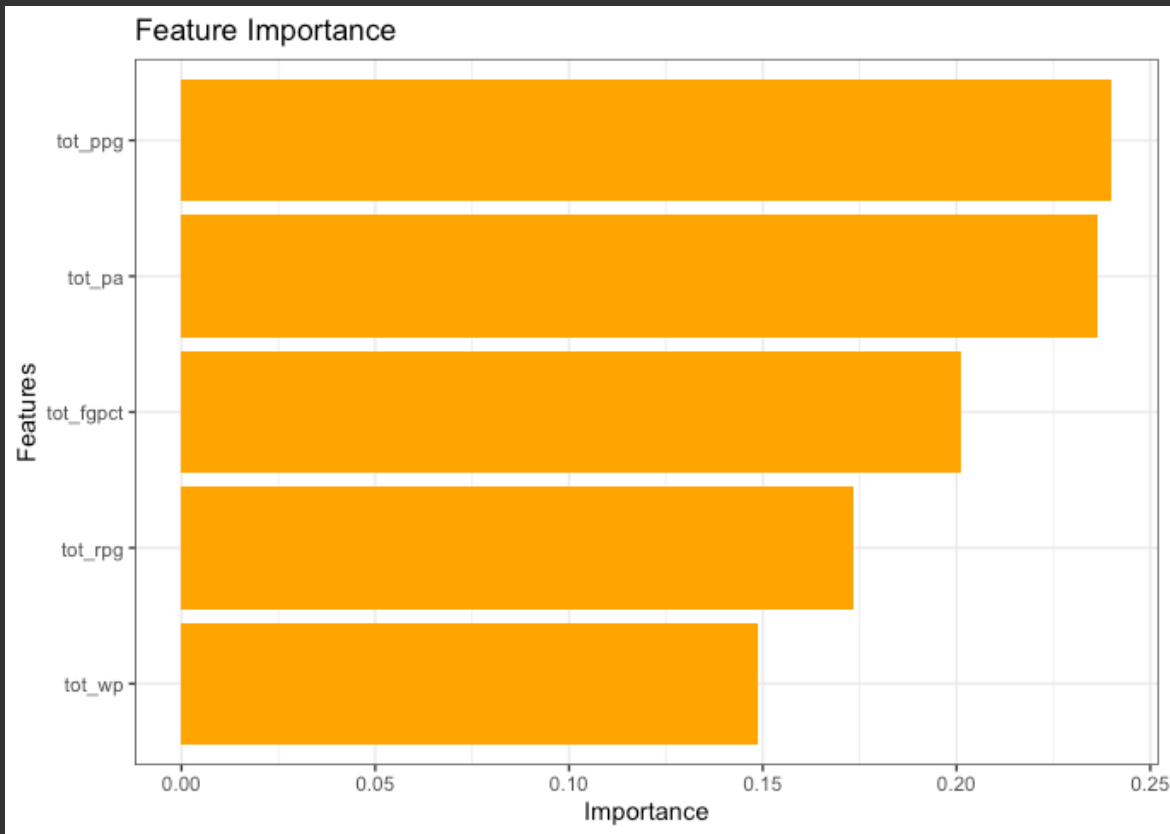
Model 6 Calibration Plot



Number of games ● 500 ● 1000 ● 1500

Predictors: tot_ppg, tot_wp, tot_fgpc, tot_pa, tot_rpg

TRAINING: XGBOOST TREE



TESTING R CODE

```
test_lin_pred <- predict(train.log_tot6,test,type = "response")
caret::confusionMatrix(as.factor(if_else(test_lin_pred>=.5,1,0)),as.factor(test$over))

xgtest <- test %>% select(tot_ppg , tot_wp , tot_fgpc , tot_pa , tot_rpg)

test_xg_pred <- predict(xg_boosted_tree1,as.matrix(xgtest), type = "response")
caret::confusionMatrix(as.factor(if_else(test_xg_pred>=.5,1,0)),as.factor(test$over))
```

Logistic Model Results

	Reference	
Prediction	0	1
0	703	736
1	328	354

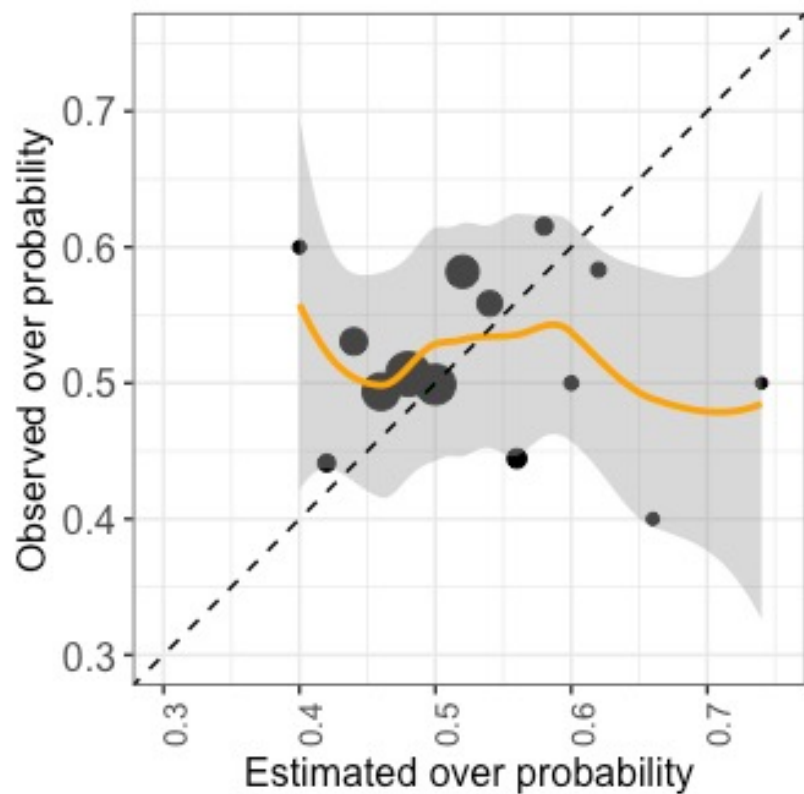
<u>Logistic</u>	<u>Xgboost</u>
Accuracy: 49.83%	Accuracy: 50.54%
Sensitivity: 68.19%	Sensitivity: 61.11%
Specificity: 32.48%	Specificity: 40.55
Detection Prevalence: 67.85%	Detection Prevalence: 60.25%
Better on unders comparatively	Better on overs comparatively

Xgboost Results

	Reference	
Prediction	0	1
0	630	648
1	401	442

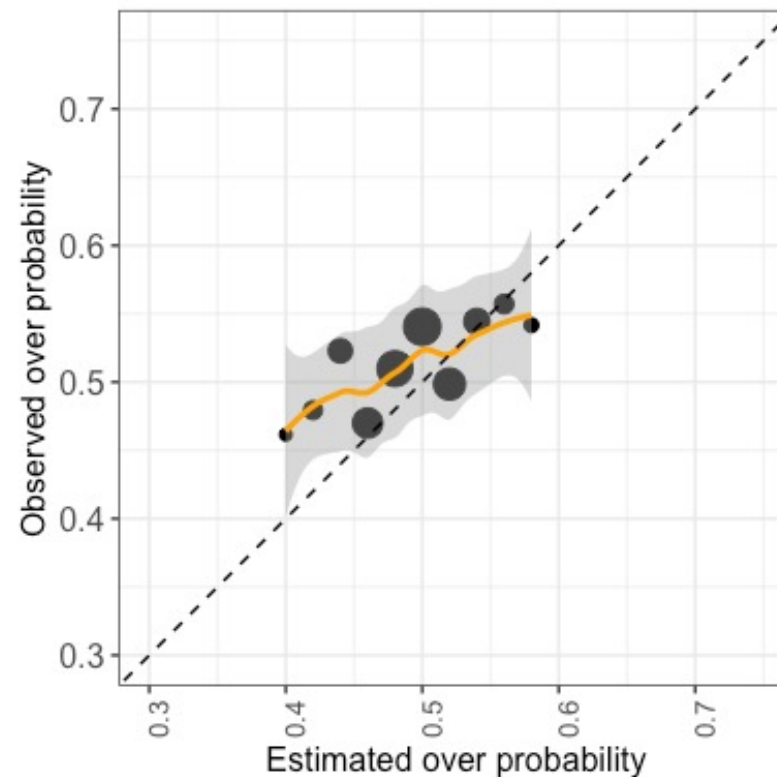
XGBOOST IS SIGNIFICANTLY MORE PRECISE

Logistic Calibration Plot



Number of games • 0 ● 100 ● 200 ● 300

XGBoost Calibration Plot



Number of games • 0 ● 50 ● 100 ● 150 ● 200 ● 250

CONCLUSIONS AND NEXT STEPS

- More robust team data
 - FGs attempted (2 and 3 point), possessions, etc....
- Odds and implied probability*

American Odds	-138
Win %	57.98%

- In-season data
 - Injuries or trades
- Moving averages
 - To show team improvement

THANK YOU