

Predict Expected Goals

- The team with more expected goals wins 65% of the time.
- Identify key offensive variables that drive play.
- Split test/train validation set at halfway point of season.
- Linear regression with stepwise selection, weighted least squares, LASSO, ridge, and tuned random forest for prediction.

Main Results

- Body checking is associated with a reduction in expected goals.
- Shots on goal and rebounds are important predictors.
- Rebounds drive high-danger shots.
- Results are for even-strength play and do not account for man-advantage play. Should investigate drivers of man-advantage offense.
- Teams can identify areas of need to improve goal-scoring ability.

NHL teams take note: hit less, shoot more, win more!

Kristina Frazier and Mike McHenry

Executive Summary

What combination of game statistics creates the strongest and most parsimonious models for expected goals scored per game and expected results in the NHL? Using data from Money puck.com we find that reducing body checking and increasing rebounds improves offensive performance, home games and high danger shots positively predict majority goals won. Hockey coaches and players should take note of these results, as they are critical in improving team performance.

Predict Majority Wins

- The target variable, xresult, indicates if a team won the majority of goals in a season, based on qualified goals for and against.
- Predictors were selected for their relationship with xresult, addressing assumptions like linearity and multicollinearity, and testing models like logistic regression.
- The final model was chosen based on classification accuracy and performance.

$$xresult = fenwickPercentage + poly(highDangerxGoalsFor, 2) + poly(highDangerxGoalsAgainst, 2) + poly(xGoalsFromxReboundsOfShotsFor, 2) + shotAttemptsFor + home\_or\_away$$

Results/Implications

- All models performed well, with the logistic regression model achieving 90% accuracy in both train/test splits and cross-validation.
- Key insights showed teams performed better at home, higher shot attempts negatively correlated with winning goals, Teams with better-quality shot attempts (higher Fenwick Percentage) are much more likely to win more goals.

Assumptions & Limitations

- The target variable measures scoring chances, not actual scoreboard results.
- Predictors showed quadratic relationships with the target, revealing non-linear dynamics like high-danger goals.

Take shots on net, not shots at others!

Table 1: Predicting Expected Goals in

Method	Prediction MSE
Stepwise Linear Regression	0.0288
Stepwise WLS Regression	0.0287
Ridge Regression	0.0310
LASSO Regression	0.0304
Random Forest	0.0472

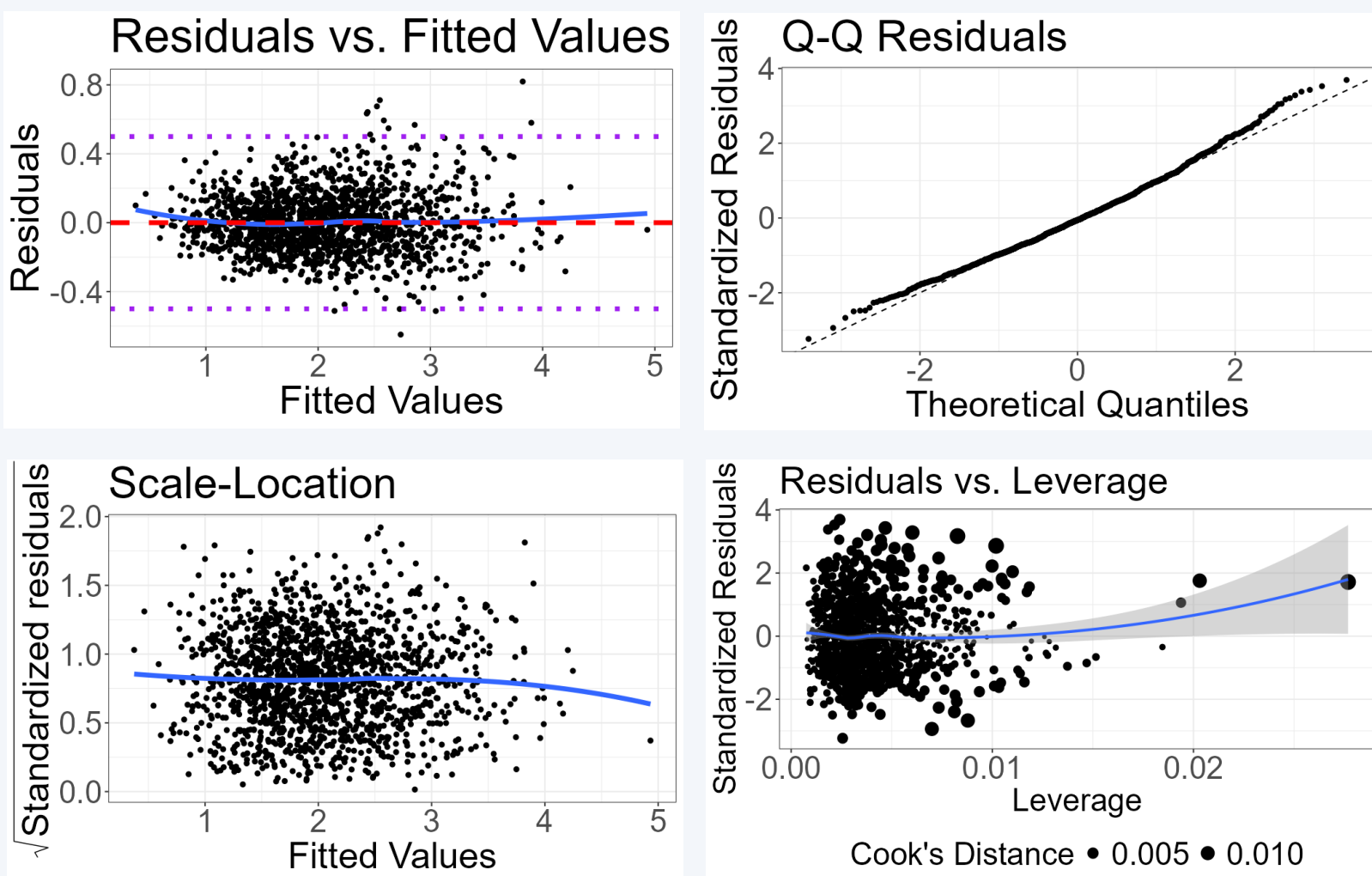
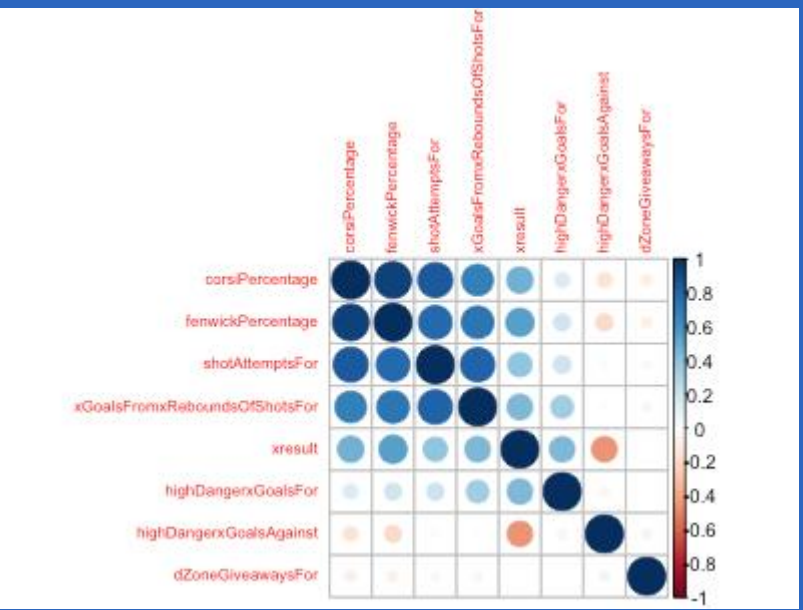
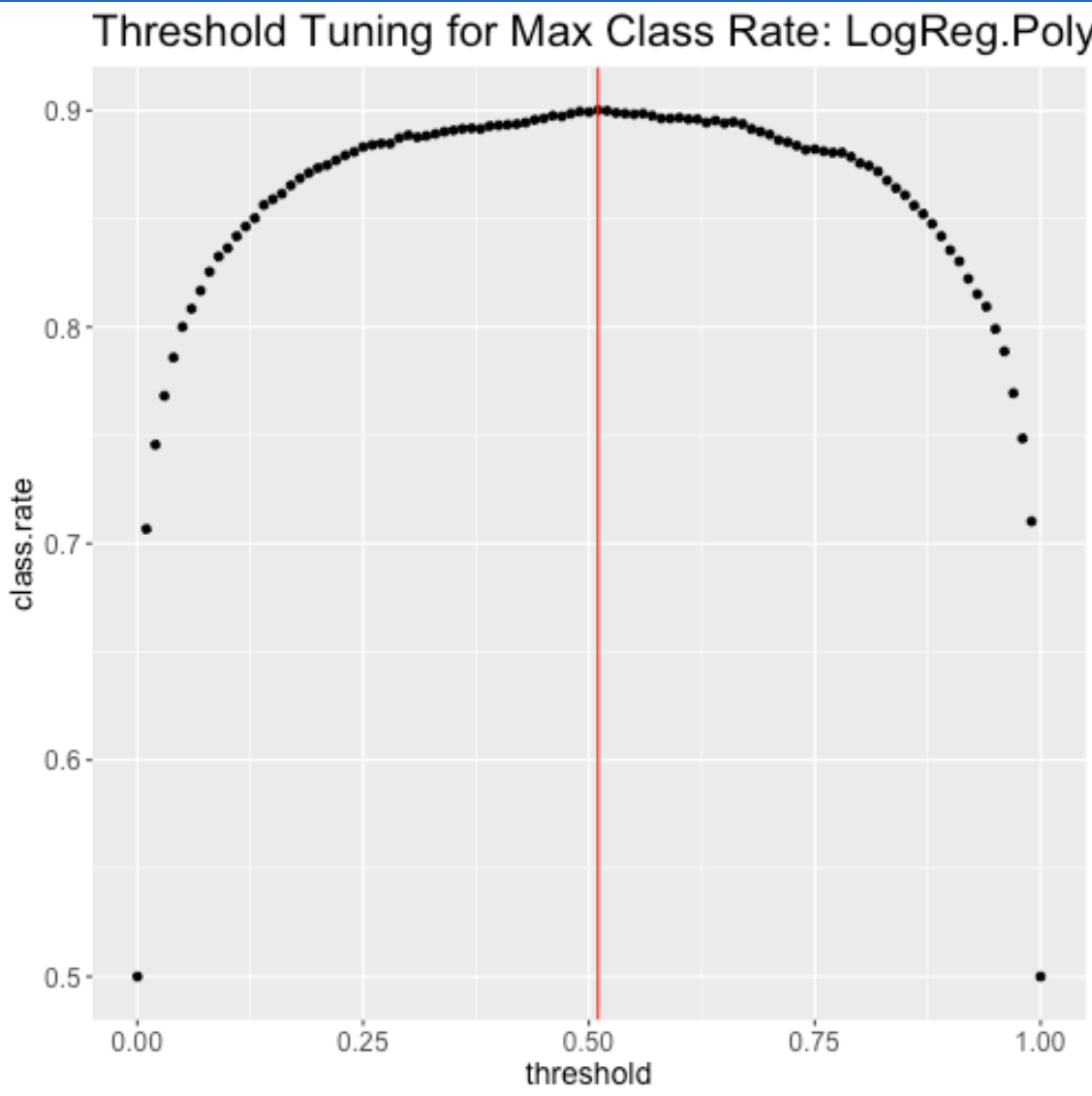
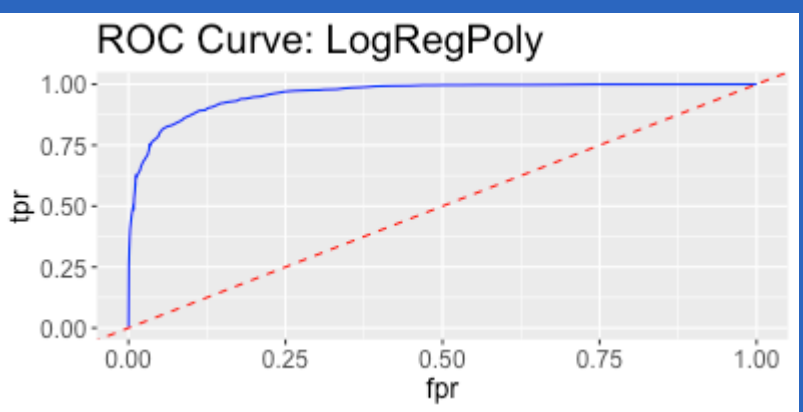
Table 2: Weighted Least Squares Results

Predictor	Estimate	Std. Error
High Danger Shots	0.2737***	(0.0035)
Medium Danger Shots	0.1257***	(0.0017)
Low Danger Shots	0.0251***	(0.0007)
Rebounds	0.0226***	(0.0034)
Hits	-0.0012*	(0.0006)

... Who nets the most? Let's classify.

Table 3: Classification Performance

Model	Accuracy
LogReg Test/Train	0.891
LogReg (Polynomial) Test/Train	0.886
LogReg K-Fold 10	0.898
LogReg (Polynomial) K-Fold 10	0.899
LDA Test/Train	0.893
LDA LOOCV	0.893
QDA Test/Train	0.888
QDA LOOCV	0.886
KNN K=42	0.841



Key References and Acknowledgments

- Gu, Wei & Saaty, Thomas. (2019). Predicting the Outcome of a Tennis Tournament: Based on Both Data and Judgments. Journal of Systems Science and Systems Engineering. 28. 317-343.
- Graham (2022). New NHL Predictive Models. The Commute Sports.
- Thanks to Professor Richard Ressler.

Data

- NHL team level data from Money puck.com for all NHL games from October 2023 through October 2024.
- 3,106 observations for even-strength play containing variables on key offensive and defensive variables.
- The dataset does not have any way of indicating wins in overtime, which means not all wins are accounted for and some games are counted as ties.