

Regression Analysis on Salary of NBA Players

Tianyi Zhou¹, Fengjie Chen², Yikun Zhang²

¹Information School, ²Departments of Statistics,
University of Washington

2020-12-14

Background

Background

- ▶ As the premier men's professional basketball league in the world, National Basketball Association (NBA) embraces high reputations in modern competitive sports and captures the eyes of millions of basketball fans worldwide.
- ▶ Big fans judge the performance of a superstar simply based on the point that he scored each game or in the whole season, while basketball mavens evaluate the capability and potential of a player by more statistics.
- ▶ Goal: Predict the salary of a NBA player in 2017 based on his season-long performance statistics.
- ▶ Evaluation metrics: We choose Mean Squared Error (MSE) as a metric to evaluate performance of a model in 5-fold cross validation.

Data Description

- ▶ Datasets available on Kaggle, skip procedure of collection.
- ▶ Combine team statistics with individual statistics, remove adundant predictors and impute missing values by our understandings of those statistics

Data Description

Table 1: Head of selected columns

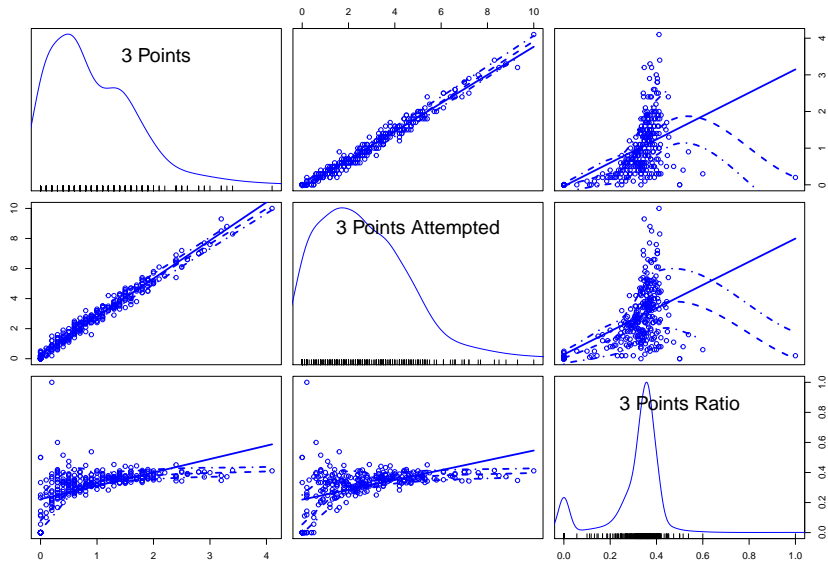
| Player | Age | Turnover | Foul | Points | Salary |
|-------------------|-----|----------|------|--------|--------|
| Russell Westbrook | 28 | 1.9 | 2.3 | 31.6 | 26.5 |
| James Harden | 27 | 1.9 | 2.7 | 29.1 | 26.5 |
| Isaiah Thomas | 27 | 1.3 | 2.2 | 28.9 | 6.6 |
| Anthony Davis | 23 | 1.2 | 2.2 | 28.0 | 22.1 |
| DeMarcus Cousins | 26 | 1.5 | 3.9 | 27.0 | 17.0 |
| Damian Lillard | 26 | 1.3 | 2.0 | 27.0 | 24.3 |
| LeBron James | 32 | 1.6 | 1.8 | 26.4 | 31.0 |
| Kawhi Leonard | 25 | 1.1 | 1.6 | 25.5 | 17.6 |

Data Preprocessing

Data Preprocessing

- ▶ Too many predictors (team, individual, social media etc.), relatively few samples, implicit relationships between predictors.
- ▶ Multicollinearity (Remove some predictors which can be determined by others, regression only on predictors with high correlation with responses, . . .)

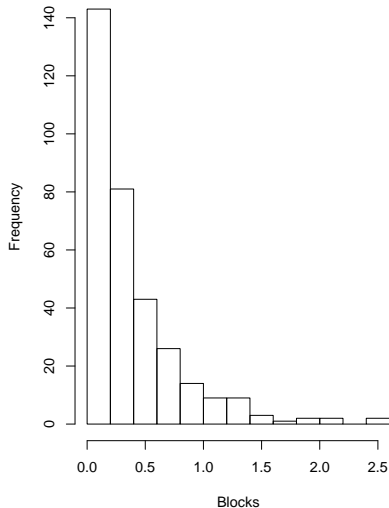
Data Preprocessing



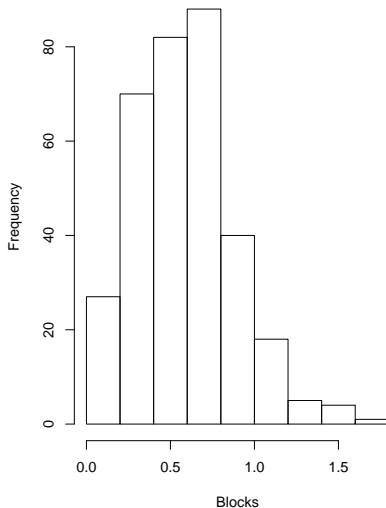
Data Preprocessing

After preprocessing, data is of shape 335×48

Blocks before transformation



Blocks after transformation



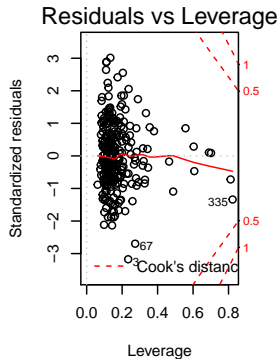
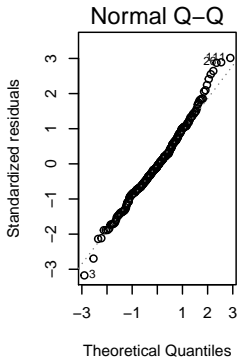
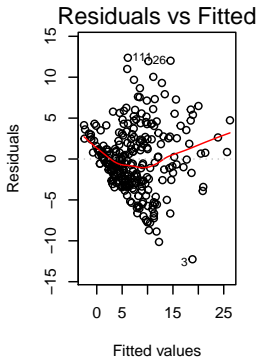
Exploratory Data Analysis

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|-------|------|------|------|------|------|--------|------|-------|------|------|------|-------|------|------|------|-------|-------|-------|------|------|------|-------|------|------|-------|-------|--------|------|------|
| AGE_x | 1 | 0.048 | 0.088 | 0.078 | 0.020 | 0.048 | 0.003 | 0.036 | 1.30 | 0.348 | 0.80 | 1.20 | 0.55 | 0.036 | 0.12 | 0.18 | 0.8 | 0.13 | 0.1 | 0.1 | 0.1000 | 0.32 | 1.035 | 0.64 | 1.1 | 0.04 | 0.1 | 0.19 | 0.18 | 0.69 | 0.48 | 0.007 | 0.707 | 1.3 | 0.14 | 0.10 | 0.799 | 0.54 | 0.11 | | | | | |
| PG | 0.04 | 1 | 0.33 | 0.050 | 0.31 | 0.90 | 0.29 | 0.84 | 3.80 | 4.1 | 0.680 | 0.690 | 0.65 | 0.48 | 0.26 | 0.10 | 0.94 | 0.92 | 0.7 | 0.7 | 0.65 | 0.7 | 0.24 | 0.52 | 0.6 | 0.47 | 0.43 | 0.23 | 0.31 | 0.18 | 0.029 | 0.08 | 0.03 | 0.4 | 2.3 | 0.38 | 0.46 | 0.66 | 0.72 | 0.80 | 0.37 | 0.4 | | |
| FG% | 0.088 | 0.33 | 1 | 0.22 | 0.18 | 0.46 | 0.65 | 0.23 | 0.04 | 0.58 | 0.4 | 0.028 | 1.1 | 0.45 | 0.19 | 0.36 | 0.22 | 0.23 | 0.17 | 0.29 | 0.27 | 0.53 | 0.26 | 0.31 | 0.18 | 0.16 | 0.34 | 0.2 | 0.16 | 0.19 | 0.83 | 0.36 | 0.4 | 0.18 | 0.7 | 0.8 | 0.4 | 0.70 | 0.8 | 0.48 | 0.13 | | | |
| 3P | 0.072 | 0.59 | 0.22 | 1 | 0.68 | 0.30 | 0.86 | 0.47 | 0.4 | 0.28 | 0.17 | 0.54 | 0.46 | 0.70 | 0.48 | 0.18 | 0.64 | 0.61 | 0.67 | 0.36 | 0.45 | 0.25 | 0.24 | 0.35 | 0.35 | 0.25 | 0.33 | 0.5 | 0.3 | 0.3 | 0.1 | 0.6 | 0.34 | 0.47 | 0.4 | 0.36 | 0.19 | 0.20 | 0.12 | 0.43 | 0.08 | 0.59 | 0.48 | 0.83 |
| 3PM | 0.023 | 0.31 | 0.16 | 0.68 | 1 | 1.08 | 0.14 | 0.13 | 0.36 | 0.052 | 0.35 | 0.24 | 1.90 | 0.21 | 0.18 | 0.32 | 0.31 | 0.3 | 0.31 | 0.17 | 0.18 | 0.67 | 0.22 | 0.13 | 0.36 | 0.29 | 0.17 | 0.25 | 0.3 | 0.19 | 0.57 | 0.3 | 0.3 | 0.36 | 0.19 | 0.22 | 0.33 | 0.06 | 0.6 | 0.7 | 0.8 | 0.89 | | |
| FT% | 0.030 | 0.94 | 0.46 | 0.30 | 0.05 | 1 | 0.8 | 0.82 | 0.57 | 0.72 | 0.62 | 0.59 | 0.54 | 0.64 | 0.87 | 0.63 | 0.57 | 0.06 | 0.10 | 0.73 | 0.17 | 0.44 | 0.57 | 0.43 | 0.36 | 0.2 | 0.28 | 0.66 | 0.4 | 0.14 | 0.23 | 0.27 | 0.73 | 0.43 | 0.64 | 0.04 | 0.34 | 0.02 | 0.14 | 0.1 | 0.1 | | | |
| 2PM | 0.06 | 0.29 | 0.85 | 0.08 | 0.14 | 0.36 | 1 | 0.2 | 0.08 | 0.42 | 0.32 | 0.15 | 14 | 0.37 | 0.13 | 0.29 | 0.19 | 0.21 | 0.2 | 0.23 | 0.27 | 0.43 | 0.16 | 0.28 | 0.11 | 0.30 | 0.29 | 0.13 | 0.80 | 0.17 | 0.33 | 0.42 | 0.3 | 0.3 | 0.12 | 0.8 | 0.7 | 0.8 | 0.58 | 0.66 | 0.12 | | | |
| FTO | 0.005 | 0.84 | 0.23 | 0.47 | 0.21 | 0.82 | 0.2 | 1 | 0.45 | 0.35 | 0.63 | 0.66 | 0.37 | 0.81 | 0.54 | 0.91 | 0.78 | 0.7 | 0.1 | 0.67 | 0.1 | 0.23 | 0.34 | 0.58 | 0.33 | 0.32 | 0.19 | 0.4 | 0.11 | 0.03 | 0.05 | 0.12 | 0.6 | 0.88 | 0.22 | 0.47 | 0.003 | 0.42 | 0.02 | 0.24 | | | | |
| FT% | 0.13 | 0.30 | 0.04 | 0.43 | 0.37 | 0.28 | 0.67 | 0.43 | 1 | 0.63 | 0.10 | 0.18 | 0.20 | 0.3 | 0.16 | 0.37 | 0.33 | 0.32 | 0.21 | 0.19 | 0.2 | 0.23 | 0.24 | 0.27 | 0.1 | 0.33 | 0.13 | 0.21 | 0.02 | 0.20 | 0.32 | 0.30 | 0.32 | 0.11 | 0.21 | 0.22 | 0.43 | 0.04 | 0.38 | 0.03 | 0.1 | 0.46 | | |
| ORB | 0.43 | 0.15 | 0.50 | 0.28 | 0.34 | 0.57 | 0.42 | 0.35 | 0.15 | 1 | 0.73 | 0.13 | 0.2 | 0.7 | 0.33 | 0.59 | 0.31 | 0.37 | 0.12 | 0.49 | 0.4 | 0.45 | 0.78 | 0.25 | 0.28 | 0.26 | 0.18 | 0.09 | 0.2 | 0.33 | 0.37 | 0.79 | 0.67 | 0.78 | 0.19 | 0.33 | 0.33 | 0.09 | 0.48 | 0.11 | 0.4 | 0.0092 | | |
| DRB | 0.06 | 0.69 | 0.41 | 0.18 | 0.02 | 0.72 | 0.32 | 0.63 | 0.11 | 0.73 | 1 | 0.38 | 0.49 | 0.68 | 0.61 | 0.66 | 0.4 | 0.68 | 0.42 | 0.3 | 0.67 | 0.68 | 0.18 | 0.38 | 0.53 | 0.35 | 0.29 | 0.1 | 0.08 | 0.08 | 0.23 | 0.3 | 0.68 | 0.6 | 0.05 | 0.32 | 0.3 | 0.33 | 0.63 | 0.6 | 0.10 | 0.39 | | |
| AST | 0.12 | 0.69 | 0.02 | 0.54 | 0.35 | 0.62 | 0.01 | 0.66 | 0.33 | 0.10 | 0.13 | 0.38 | 1 | 0.71 | 0.08 | 0.84 | 0.36 | 0.71 | 0.73 | 0.65 | 0.1 | 0.57 | 0.52 | 0.16 | 0.37 | 0.47 | 0.27 | 0.33 | 0.12 | 0.84 | 0.61 | 0.66 | 0.5 | 0.18 | 0.28 | 0.08 | 0.68 | 0.15 | 0.14 | 0.03 | 0.58 | 0.65 | | |
| STL | 0.05 | 0.65 | 0.11 | 0.46 | 0.24 | 0.52 | 0.14 | 0.6 | 0.18 | 0.2 | 0.48 | 0.71 | 0.31 | 0.68 | 0.49 | 0.65 | 0.74 | 0.59 | 0.17 | 0.67 | 0.46 | 0.2 | 0.41 | 0.44 | 0.27 | 0.33 | 0.027 | 0.45 | 0.33 | 0.3 | 0.2 | 0.05 | 0.1 | 0.44 | 0.16 | 0.24 | 0.34 | 0.58 | 0.66 | 0.50 | 0.11 | | | |
| BLK | 0.05 | 0.34 | 0.37 | 0.18 | 0.71 | 0.54 | 0.37 | 0.5 | 0.03 | 0.7 | 0.65 | 0.83 | 0.31 | 1 | 0.3 | 0.65 | 0.37 | 0.45 | 0.16 | 0.52 | 0.46 | 0.43 | 0.34 | 0.31 | 0.3 | 0.24 | 0.67 | 0.18 | 0.23 | 0.44 | 0.33 | 0.52 | 0.1 | 0.34 | 0.35 | 0.13 | 0.78 | 0.68 | 0.38 | 0.59 | | | | |
| TOK | 0.03 | 0.82 | 0.19 | 0.48 | 0.1 | 0.8 | 0.18 | 0.81 | 0.1 | 0.63 | 0.83 | 0.6 | 0.56 | 0.02 | 0.59 | 0.6 | 0.25 | 0.36 | 0.53 | 0.7 | 0.28 | 0.18 | 0.58 | 0.22 | 0.16 | 0.28 | 0.58 | 0.22 | 0.16 | 0.28 | 0.58 | 0.22 | 0.16 | 0.28 | 0.58 | 0.22 | 0.16 | 0.28 | 0.58 | 0.22 | 0.16 | 0.28 | 0.58 | |
| PF | 0.01 | 0.61 | 0.35 | 0.18 | 0.10 | 0.66 | 0.29 | 0.54 | 0.16 | 0.59 | 0.65 | 0.39 | 0.49 | 0.65 | 0.63 | 1 | 0.52 | 0.66 | 0.28 | 0.31 | 0.46 | 0.33 | 0.18 | 0.37 | 0.33 | 0.4 | 0.21 | 0.23 | 0.03 | 0.88 | 0.58 | 0.21 | 0.32 | 0.31 | 0.21 | 0.3 | 0.3 | 0.2 | 0.95 | 0.68 | 0.23 | 0.3 | | |
| MFG | 0.08 | 0.94 | 0.22 | 0.62 | 0.32 | 0.67 | 0.19 | 0.91 | 0.33 | 0.31 | 0.64 | 0.17 | 0.65 | 0.37 | 0.83 | 0.52 | 1 | 0.86 | 0.5 | 0.06 | 0.74 | 0.22 | 0.44 | 0.63 | 0.34 | 0.4 | 0.17 | 0.41 | 0.15 | 0.02 | 0.18 | 0.68 | 0.06 | 0.19 | 0.28 | 0.4 | 0.74 | 0.48 | 0.38 | 0.3 | 0.067 | | | |
| PTS | 0.05 | 0.92 | 0.24 | 0.30 | 0.71 | 0.83 | 0.21 | 0.78 | 0.33 | 0.7 | 0.68 | 0.73 | 0.4 | 0.3 | 0.66 | 0.6 | 0.65 | 0.61 | 0.68 | 0.52 | 0.21 | 0.53 | 0.59 | 0.44 | 0.21 | 0.3 | 0.24 | 0.1 | 0.18 | 0.00 | 0.06 | 0.16 | 0.31 | 0.39 | 0.14 | 0.68 | 0.56 | 0.13 | 0.6 | | | | | |
| ORPM | 0.15 | 0.7 | 0.17 | 0.67 | 0.36 | 0.57 | 0.2 | 0.7 | 0.32 | 0.12 | 0.42 | 0.65 | 0.59 | 0.16 | 0.58 | 0.28 | 0.8 | 0.65 | 1 | 0.05 | 0.61 | 0.65 | 0.18 | 0.51 | 0.5 | 0.17 | 0.55 | 0.05 | 0.46 | 0.35 | 0.21 | 0.18 | 0.48 | 0.52 | 0.28 | 0.33 | 0.44 | 0.59 | 0.68 | 0.48 | 0.6 | 0.23 | | |
| DRPM | 0.16 | 0.47 | 0.25 | 0.36 | 0.31 | 0.55 | 0.23 | 0.72 | 0.2 | 0.49 | 0.45 | 0.14 | 0.17 | 0.53 | 0.2 | 0.43 | 0.08 | 0.1 | 0.08 | 1 | 0.11 | 0.17 | 0.7 | 0.1 | 0.09 | 0.18 | 0.62 | 0.3 | 0.19 | 0.18 | 0.06 | 0.46 | 0.53 | 0.58 | 0.24 | 0.20 | 0.11 | 0.22 | 0.08 | 0.7 | 0.02 | 0.2 | | |
| WINS_RPM | 0.1 | 0.65 | 0.27 | 0.40 | 0.17 | 0.61 | 0.27 | 0.67 | 0.19 | 0.4 | 0.67 | 0.57 | 0.67 | 0.48 | 0.59 | 0.46 | 0.74 | 0.68 | 0.81 | 0.4 | 1 | 0.68 | 0.15 | 0.75 | 0.54 | 0.15 | 0.46 | 0.35 | 0.21 | 0.33 | 0.12 | 0.22 | 0.18 | 0.32 | 0.41 | 0.14 | 0.09 | 0.03 | 0.38 | 0.31 | | | | |
| PIE_x | 0.11 | 0.7 | 0.53 | 0.25 | 0.16 | 0.73 | 0.43 | 0.71 | 0.2 | 0.48 | 0.68 | 0.52 | 0.46 | 0.43 | 0.6 | 0.33 | 0.72 | 0.52 | 0.65 | 0.7 | 0.66 | 1 | 0.11 | 0.38 | 0.51 | 0.19 | 0.47 | 0.03 | 0.42 | 0.12 | 0.08 | 0.24 | 0.44 | 0.47 | 0.12 | 0.5 | 0.58 | 0.62 | 0.78 | 0.58 | 0.45 | 0.11 | | |
| FACE | 0.000 | 0.02 | 0.02 | 0.20 | 0.16 | 0.23 | 0.20 | 0.17 | 0.60 | 0.16 | 0.4 | 0.13 | 0.25 | 0.18 | 0.22 | 0.10 | 0.78 | 0.15 | 1 | 0.09 | 0.09 | 0.29 | 0.11 | 0.2 | 0.28 | 0.29 | 0.30 | 0.08 | 0.02 | 0.00 | 0.33 | 0.1 | 0.18 | 0.05 | 0.18 | 0.08 | 0.24 | 0.15 | | | | | | |
| W_x | 0.12 | 0.52 | 0.31 | 0.34 | 0.60 | 0.42 | 0.28 | 0.34 | 0.23 | 0.25 | 0.38 | 0.73 | 0.41 | 0.34 | 0.36 | 0.37 | 0.44 | 0.53 | 0.59 | 1 | 0.57 | 0.3 | 0.09 | 1 | 0.37 | 0.26 | 0.32 | 0.28 | 0.18 | 0.08 | 0.48 | 0.33 | 0.22 | 0.16 | 0.4 | 0.41 | 0.13 | 0.25 | 0.22 | 0.68 | | | | |
| SALARY_MILLIONS | 0.35 | 0.6 | 0.18 | 0.35 | 0.13 | 0.57 | 0.11 | 0.58 | 0.24 | 0.28 | 0.53 | 0.79 | 0.44 | 0.31 | 0.53 | 0.33 | 0.63 | 0.59 | 0.5 | 0.99 | 0.54 | 0.51 | 0.98 | 0.3 | 1 | 0.16 | 0.28 | 0.11 | 0.27 | 0.13 | 0.04 | 0.12 | 0.16 | 0.11 | 0.12 | 0.29 | 0.26 | 0.41 | 0.13 | 0.12 | 0.67 | 0.1 | | |
| L_0 | 0.46 | 0.17 | 0.16 | 0.25 | 0.16 | 0.43 | 0.13 | 0.33 | 0.27 | 0.26 | 0.35 | 0.27 | 0.72 | 0.3 | 0.37 | 0.44 | 0.34 | 0.45 | 0.17 | 0.19 | 0.15 | 0.29 | 0.26 | 0.16 | 1 | 0.09 | 0.4 | 0.12 | 0.28 | 0.59 | 0.48 | 0.02 | 0.3 | 0.8 | 0.2 | 0.23 | 0.06 | 0.30 | 0.18 | 0.11 | 0.42 | | | |
| OFFRTG | 0.11 | 0.43 | 0.34 | 0.87 | 0.30 | 0.29 | 0.32 | 0.1 | 0.18 | 0.29 | 0.33 | 0.39 | 0.24 | 0.28 | 0.21 | 0.4 | 0.4 | 0.55 | 0.68 | 0.48 | 0.7 | 0.11 | 0.52 | 0.29 | 0.08 | 1 | 0.09 | 0.17 | 0.21 | 0.11 | 0.03 | 0.07 | 0.33 | 0.21 | 0.46 | 0.45 | 0.18 | 0.09 | 0.45 | 0.14 | 0.37 | | | |
| DEFRTG | 0.40 | 0.23 | 0.2 | 0.15 | 0.17 | 0.2 | 0.13 | 0.19 | 0.33 | 0.09 | 0.1 | 0.12 | 0.28 | 0.67 | 0.18 | 0.2 | 0.17 | 0.21 | 0.05 | 0.3 | 0.08 | 0.3 | 0.20 | 0.28 | 0.11 | 0.4 | 0.08 | 1 | 0.03 | 0.28 | 0.38 | 0.33 | 0.05 | 0.10 | 0.1 | 0.26 | 0.28 | 0.14 | 0.09 | 0.62 | 0.12 | 0.28 | | |
| AST% | 0.1 | 0.31 | 0.16 | 0.3 | 0.25 | 0.29 | 0.18 | 0.11 | 0.13 | 0.2 | 0.08 | 0.84 | 0.43 | 0.16 | 0.53 | 0.03 | 0.41 | 0.48 | 0.19 | 0.35 | 0.42 | 0.08 | 0.98 | 0.27 | 0.18 | 0.17 | 0.3 | 1 | 0.6 | 0.82 | 0.37 | 0.2 | 0.28 | 0.42 | 0.16 | 0.14 | 0.14 | 0.22 | 0.02 | 0.27 | 0.4 | | | |
| ASTTO | 0.19 | 0.16 | 0.18 | 0.3 | 0.03 | 0.66 | 0.17 | 0.11 | 0.21 | 0.33 | 0.09 | 0.33 | 0.23 | 0.30 | 0.17 | 0.08 | 0.15 | 0.24 | 0.35 | 0.18 | 0.21 | 0.12 | 0.28 | 0.18 | 0.13 | 0.23 | 0.21 | 0.2 | 0.6 | 1 | 0.72 | 0.48 | 0.38 | 0.44 | 0.08 | 0.58 | 0.54 | 0.43 | 0.53 | 0.1 | | | | |
| AST_RATIO | 0.18 | 0.02 | 0.18 | 0.1 | 0.19 | 0.04 | 0.18 | 0.03 | 0.22 | 0.32 | 0.12 | 0.64 | 0.3 | 0.27 | 0.28 | 0.58 | 0.2 | 0.1 | 0.21 | 0.08 | 0.13 | 0.08 | 0.30 | 0.08 | 0.48 | 0.59 | 0.11 | 0.03 | 0.82 | 0.72 | 1 | 0.4 | 0.3 | 0.37 | 0.14 | 0.18 | 0.18 | 0.00 | 0.80 | 0.68 | 0.92 | | | |
| ORB% | 0.069 | 0.88 | 0.53 | 0.6 | 0.5 | 0.14 | 0.34 | 0.05 | 0.30 | 0.79 | 0.33 | 0.3 | 0.24 | 0.09 | 0.21 | 0.13 | 0.18 | 0.47 | 0.03 | 0.24 | 0.08 | 0.48 | 0.18 | 0.48 | 0.34 | 0.33 | 0.73 | 0.48 | 0.4 | 1 | 0.67 | 0.87 | 0.33 | 0.23 | 0.18 | 0.08 | 0.00 | 0.73 | 0.42 | 0.078 | | | | |
| ORB% | 0.042 | 0.85 | 0.3 | 0.34 | 0.33 | 0.23 | 0.23 | 0.12 | 0.67 | 0.68 | 0.18 | 0.55 | 0.083 | 0.32 | 0.085 | 0.12 | 0.04 | 0.53 | 0.22 | 0.42 | 0.33 | 0.18 | 0.87 | 0.07 | 0.75 | 0.44 | 0.2 | 0.38 | 0.3 | 0.67 | 1 | 0.95 | 0.24 | 0.17 | 0.19 | 0.94 | 0.18 | 0.00 | 0.80 | 0. | | | | |

Linear Regression

Linear Regression

- ▶ SALARY ~ 42 continuous + 1 categorical with 6 levels (position)
- ▶ Leave one fold out, and train on other four folds



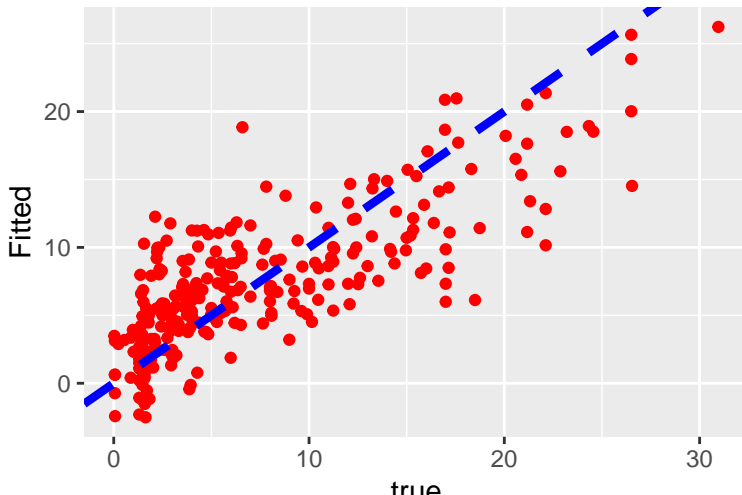
Linear Regression

- ▶ The normal distribution seems plausible, and few outliers are presented.
- ▶ However, strong evidence is presented against constant-variance assumption, inference may not be solid.

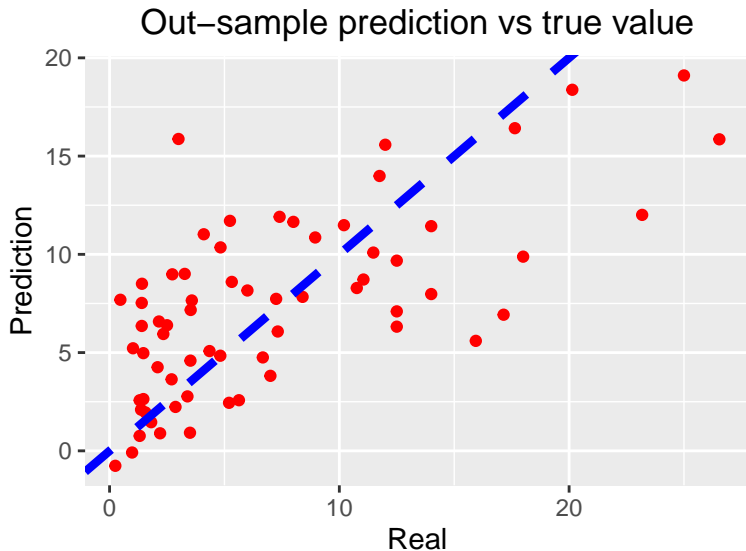
Linear Regression

- conservative on the right-hand side, negative fitted value on the left-hand side. Corrected prediction or other models may be needed.

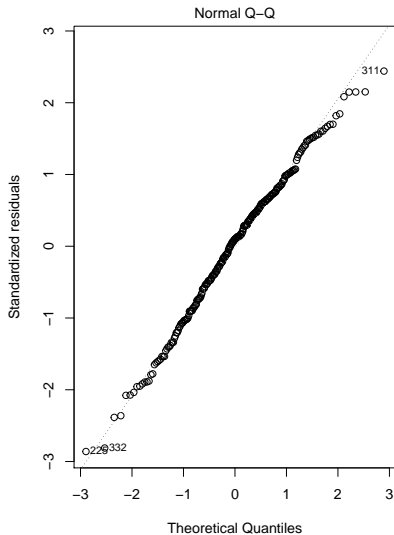
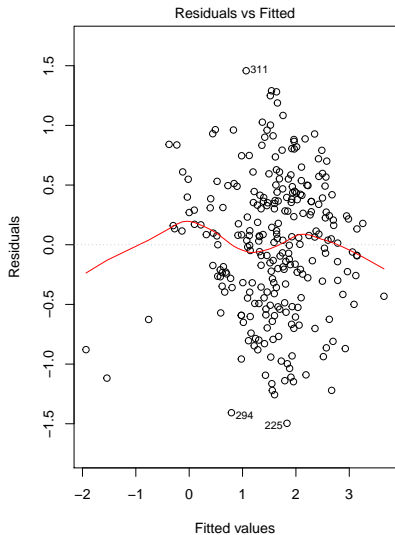
In-sample fitted vs true value



Linear Regression



Log Transformation of Response



Log Transformation of Response

- ▶ Age, 2 point%, Foul, Win, Offensive rating, Team value are significant at 5% level.
- ▶ Compared with null model, error sum of square shrinkages by 47%.

Linear Regression Results

- ▶ Linear model including all predictors: $MSE = 23.4$
- ▶ Correct the predicted value using $\max(0, x)$, $MSE = 23.2$
- ▶ After log-scaling response, $MSE = 23.8$

Regularized Linear Models

Regularized Linear Models

Lasso Regression:

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda ||\beta||_1$$

Ridge Regression:

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda ||\beta||_2^2$$

- ▶ Lasso Regression: $\text{MSE} = 27.26$
- ▶ Ridge Regression: $\text{MSE} = 32.24$

Variants of Lasso (I)

Lasso (L^1 Penalty):

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \sum_i |\beta_i|$$

$$\equiv \min_{\beta} \|y - X\beta\|_2^2 + f_{\lambda}(\beta)$$

Variants of Lasso (I)

Lasso (L^1 Penalty):

$$\begin{aligned} \min_{\beta} ||y - X\beta||_2^2 + \lambda \sum_i |\beta_i| \\ \equiv \min_{\beta} ||y - X\beta||_2^2 + f_{\lambda}(\beta) \end{aligned}$$

SCAD Penalty:

$$\min_{\beta} ||y - X\beta||_2^2 + n \sum_i p_{\lambda}(\beta_i; a),$$

where for some $a > 2$ and $\lambda > 0$

$$p_{\lambda}(\beta_i; a) = \begin{cases} \lambda|\beta_i|, & \text{if } |\beta_i| \leq \lambda, \\ -\frac{(\beta_i^2 - 2a\lambda|\beta_i| + \lambda^2)}{2(a-1)} & \text{if } \lambda < |\beta_i| \leq a\lambda, \\ (a+1)\lambda^2/2 & \text{if } |\beta_i| > a\lambda. \end{cases}$$

Variants of Lasso (II)

Lasso (Penalized or Lagrangian Form):

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda ||\beta||_1$$

Variants of Lasso (II)

Lasso (Penalized or Lagrangian Form):

$$\min_{\beta} ||y - X\beta||_2^2 + \lambda ||\beta||_1$$

\Rightarrow

$$\min_{\beta} ||y - X\beta||_2^2 \quad \text{subject to } ||\beta||_1 \leq s$$

Variants of Lasso (II)

Lasso (Penalized or Lagrangian Form):

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

\Rightarrow

$$\min_{\beta} \|y - X\beta\|_2^2 \quad \text{subject to } \|\beta\|_1 \leq s$$

Dantzig Selector:

$$\min_{\beta} \|\beta\|_1 \quad \text{subject to } \|X^T(y - X\beta)\|_{\infty} \leq s,$$

where s is a tuning parameter.

Comparison of Models

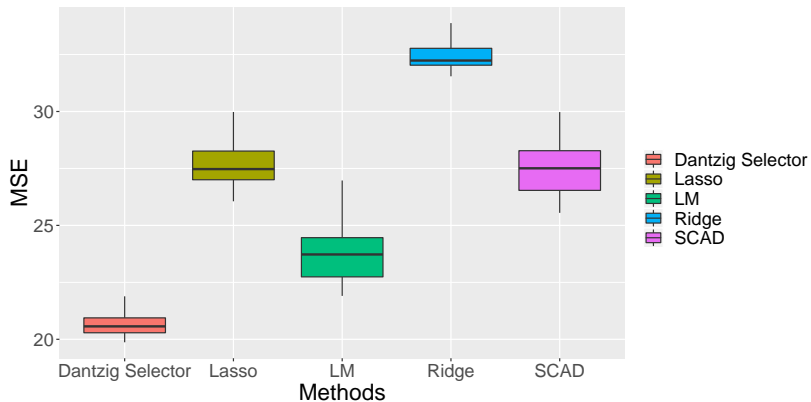


Figure 2: MSE of Different Models

Conclusions and Future Works

Conclusions and Future Works

Conclusions:

- ▶ Linear models seem to be reasonable and have relatively good predictability on the current dataset. (Compared with the null model, MSE decreases by 54%)
- ▶ Dantzig selector can further reduce MSE

Conclusions and Future Works

Conclusions:

- ▶ Linear models seem to be reasonable and have relatively good predictability on the current dataset. (Compared with the null model, MSE decreases by 54%)
- ▶ Dantzig selector can further reduce MSE

Future Works:

- ▶ Garner more predictors (e.g. social media stat), and collect more samples (from different years)
- ▶ Implement Grouped Lasso when more categorical variables are present

Thank you!