

Predicting NBA Player Salary

Using Linear Regression



Michael Cantrell, PhD

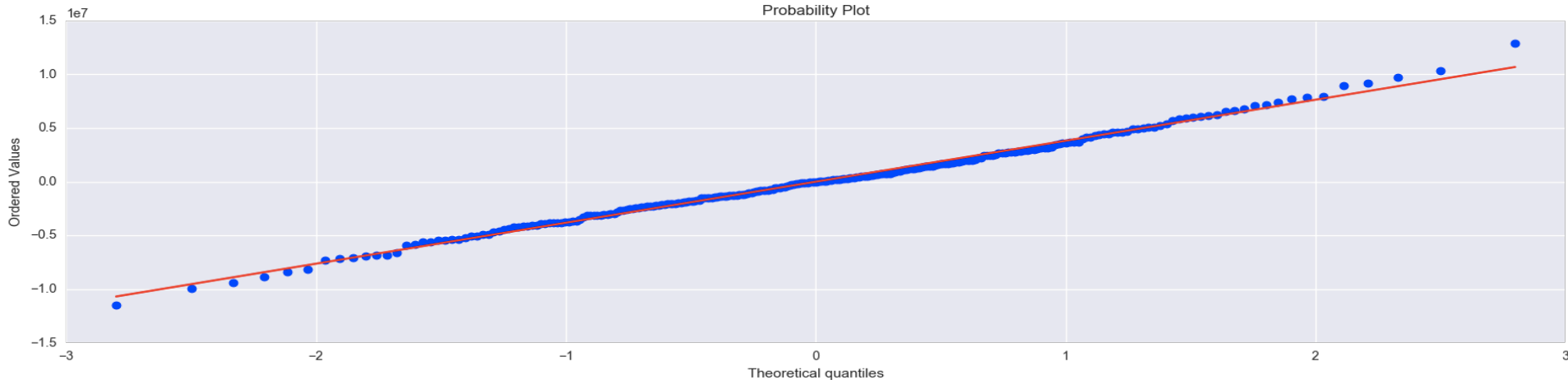
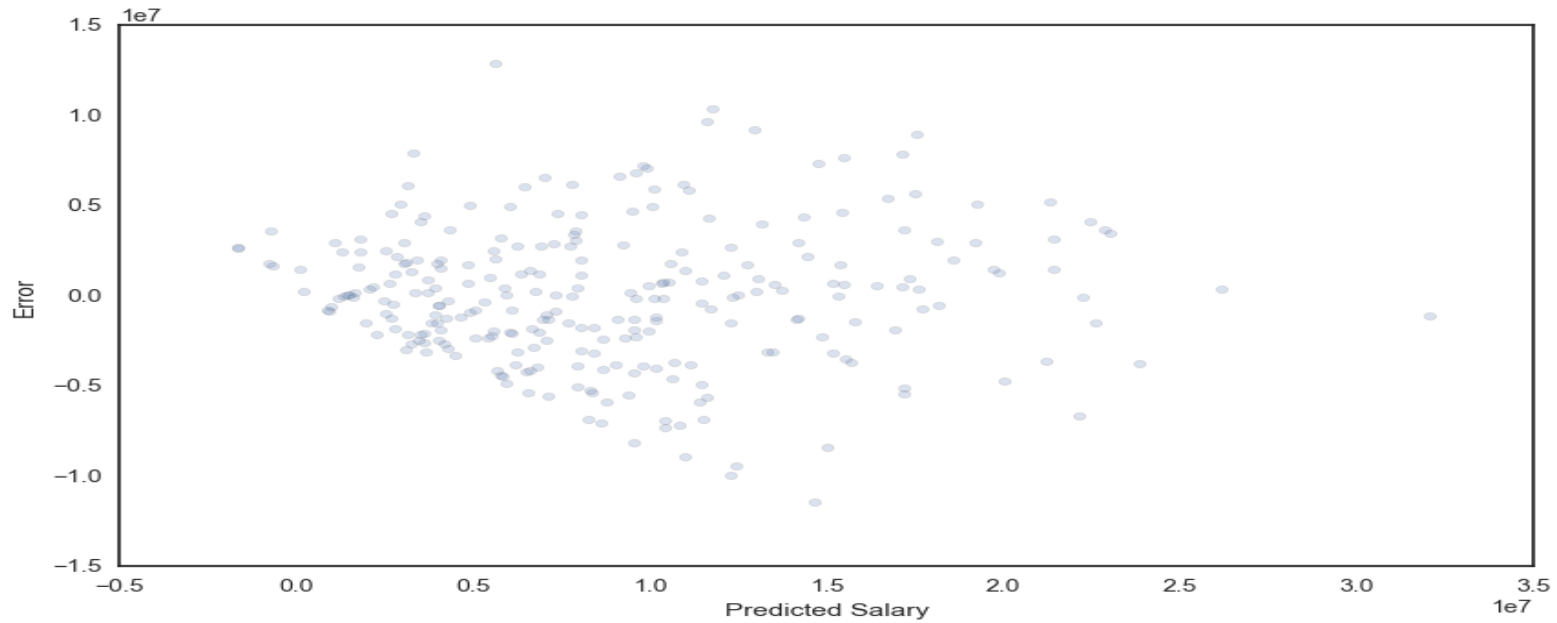
Problem

- Can future player salary be predicted by past player performance?
- What should an aspiring basketball player do?

Feature Pool

- **Age** -- Age of Player
- **Tm** -- Team
- **Pos** -- Position
- **G** -- Games Played
- **GS** -- Games Started
- **MP** -- Minutes Played
- **FG** -- Field Goals
- **FGA** -- Field Goal Attempts
- **FGP** -- Field Goal Percentage
- **3P** -- 3-Point Field Goals
- **PA3** -- 3-Point Field Goal Attempts
- **PP3** -- 3-Point Field Goal Percentage
- **2P** -- 2-Point Field Goals
- **2PA** -- 2-Point Field Goal Attempts
- **PP2** -- 2-Point Field Goal Percentage
- **eFGP** -- Effective Field Goal Percentage
- **FT** -- Free Throws
- **FTA** -- Free Throw Attempts
- **FTP** -- Free Throw Percentage
- **ORB** -- Offensive Rebounds
- **DRB** -- Defensive Rebounds
- **TRB** -- Total Rebounds
- **AST** -- Assists
- **STL** -- Steals
- **BLK** -- Blocks
- **TOV** -- Turnovers
- **PF** -- Personal Fouls
- **PTS** -- Points

Linear Regression is Appropriate

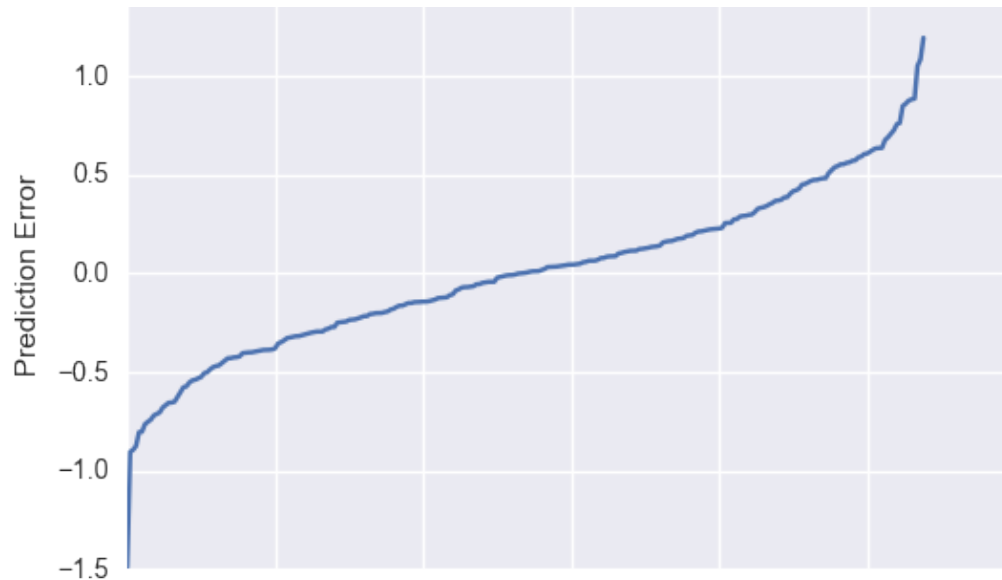


All Significant Features $R^2 = 0.5$

Predicted Salary



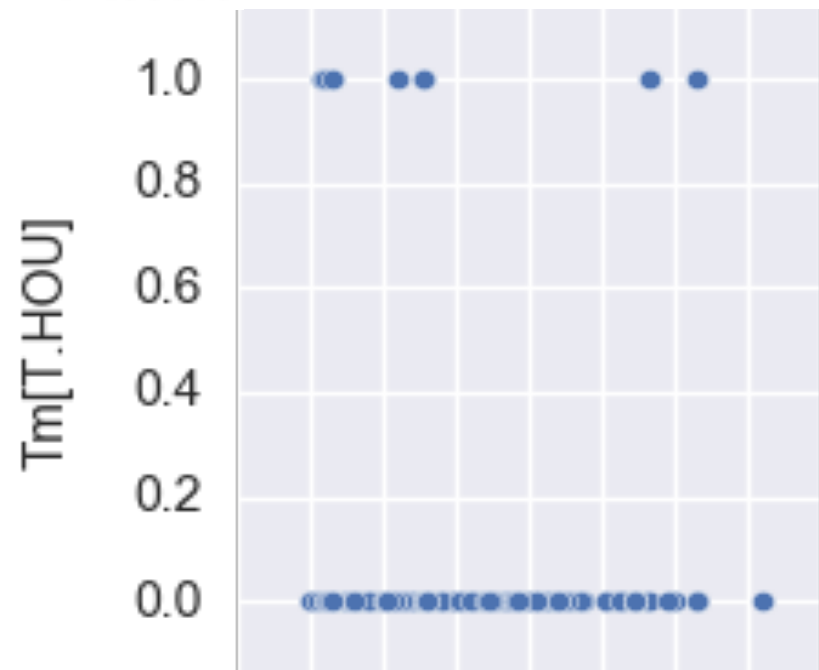
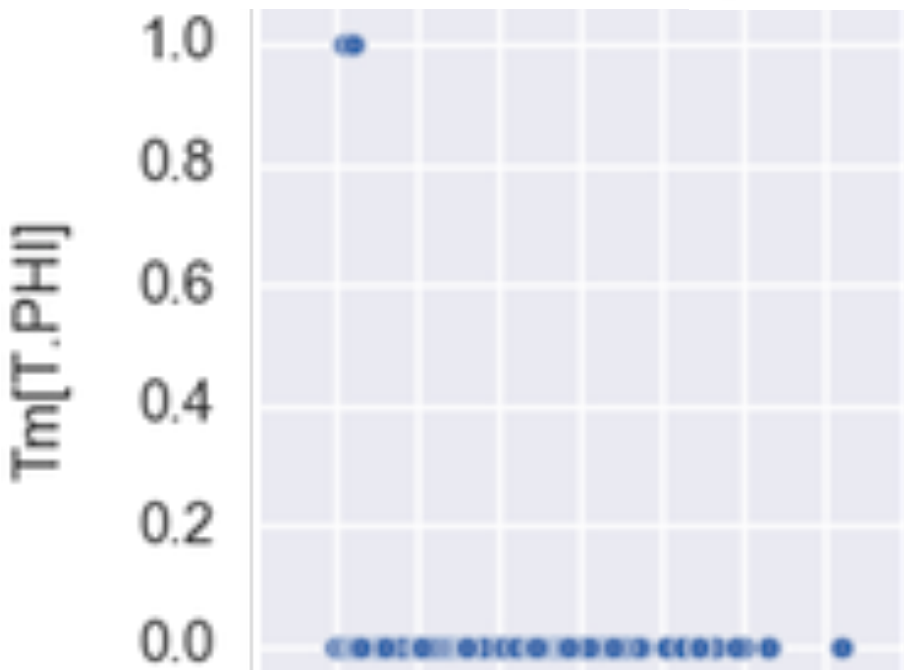
Predictive Error

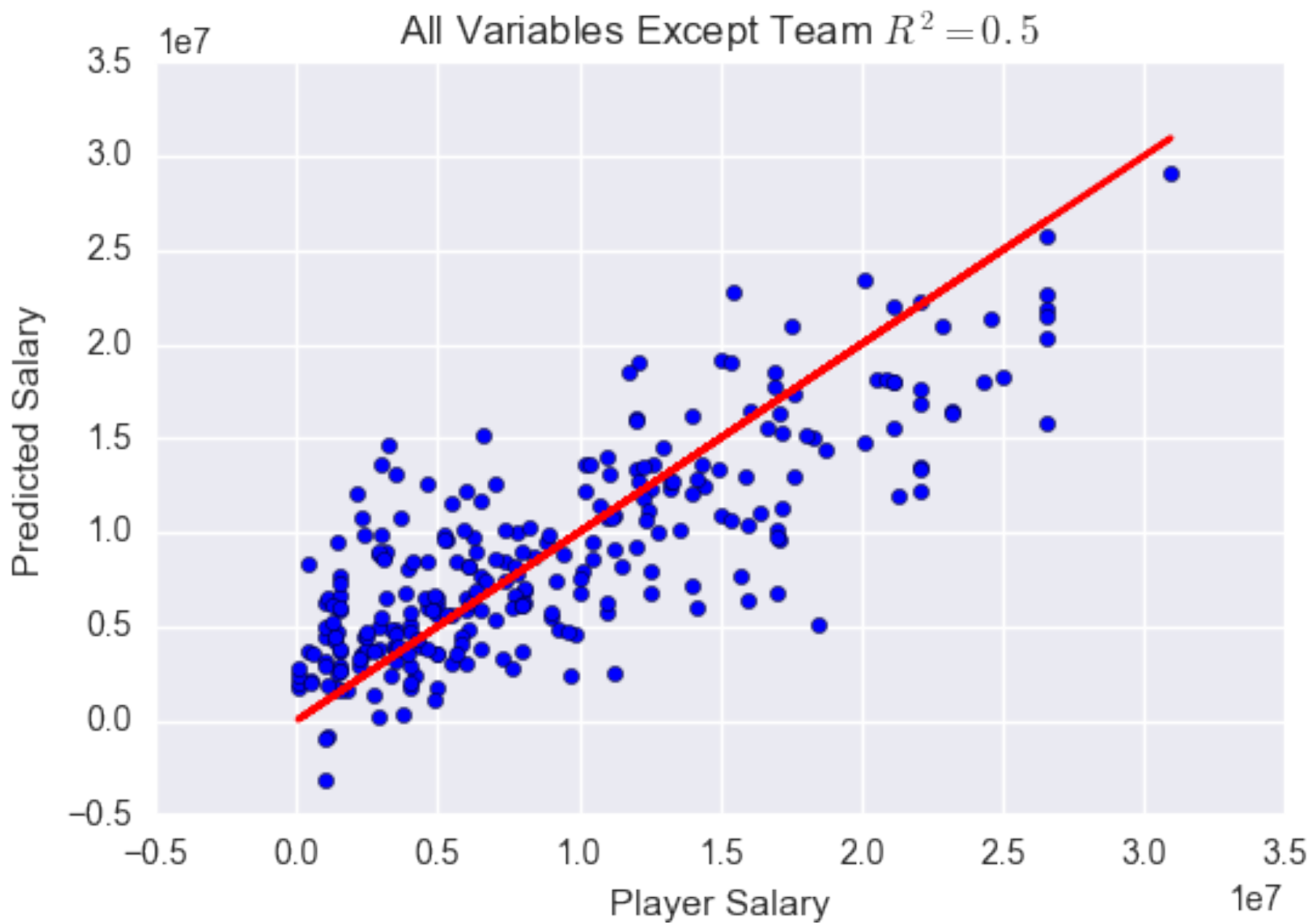


	Salary	Predicted Salary	Error		Salary	Predicted Salary	Error
Underpaid Players				Overpaid Players			
C.J. McCollum	3219579	18118572	-14898993	Harrison Barnes	22116750	13306157	8810592
Marcus Morris	4625000	13675075	-9050075	Dirk Nowitzki	25000000	16154748	8845251
Mason Plumlee	2328530	11267671	-8939141	Dwight Howard	23180275	12676165	10504109
Giannis Antetokounmpo	2995421	11735187	-8739766	Allen Crabbe	18500000	7662556	10837443
Steven Adams	3140517	11185987	-8045470	Mike Conley	26540100	14614082	11926017

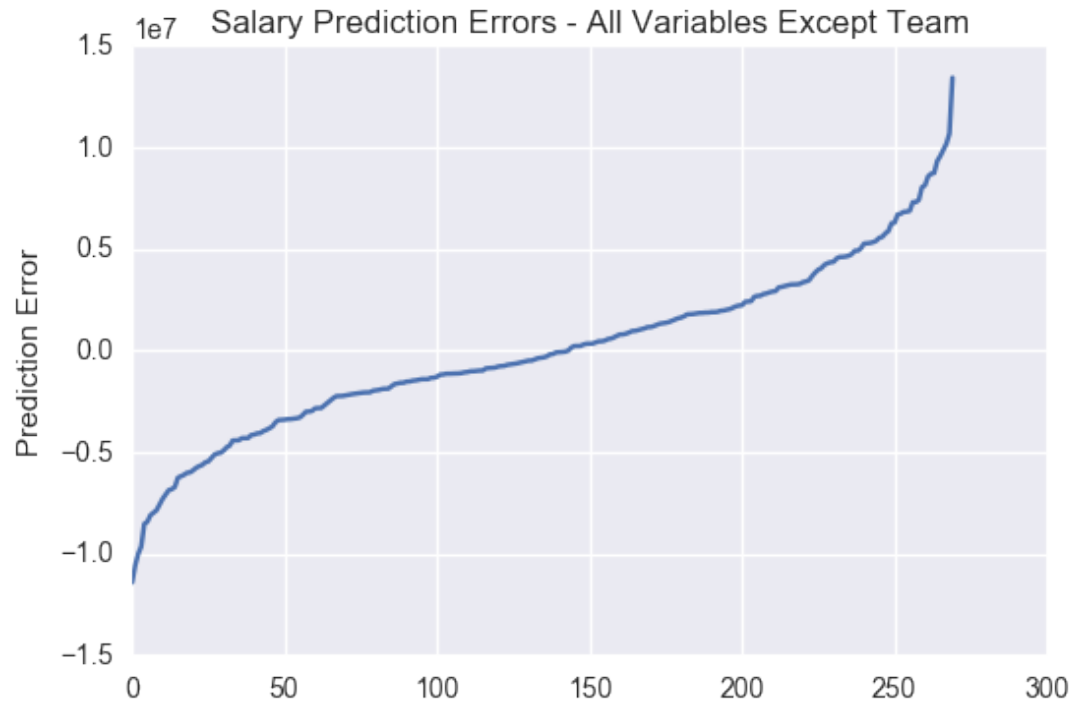
Is Team Really Predictive?

AST	0.000054
Tm[T.PHI]	0.007863
TOV	0.009825
Pos[T.PG]	0.010337
STL	0.023016
MP	0.024541
PP2	0.046080
PA3	0.062949
Tm[T.DEN]	0.064821
Tm[T.LAL]	0.081323
FG	0.095608



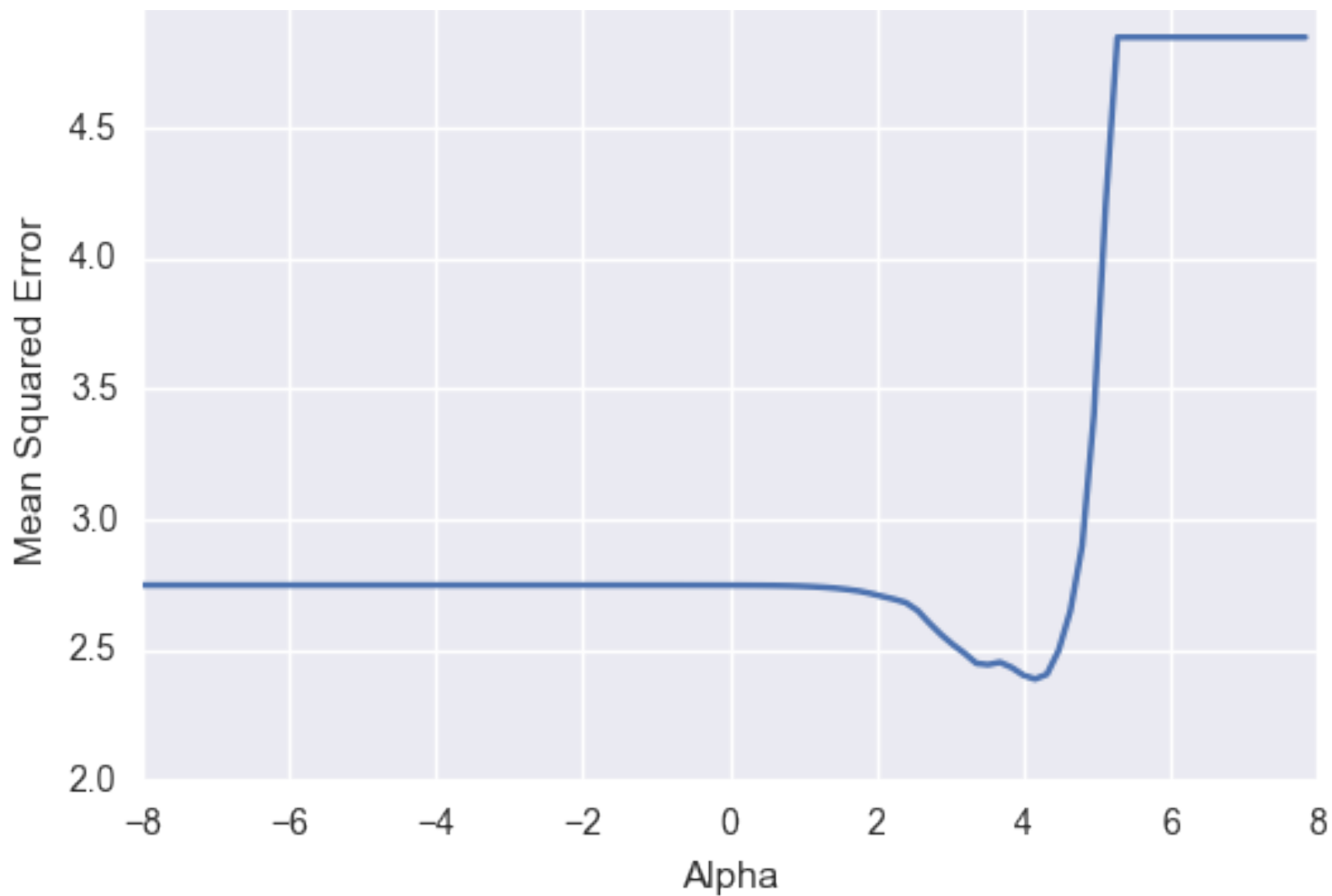


Predictive Error

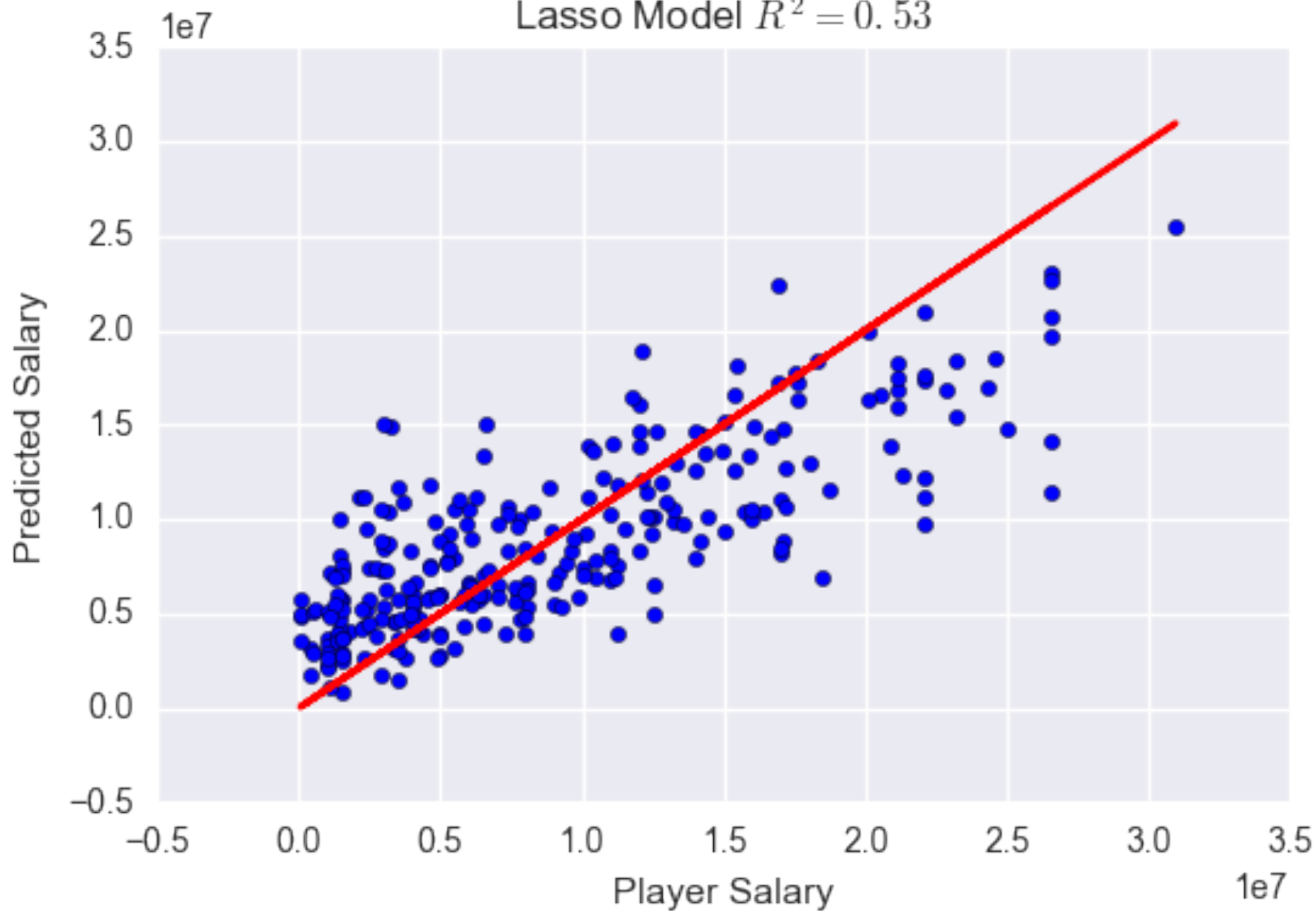


	Salary	Predicted Salary - Model 2	Error		Salary	Predicted Salary - Model 2	Error
Underpaid Players				Overpaid Players			
C.J. McCollum	3219579	14622930	-11403351	Timofey Mozgov	16000000	6429188	9570811
Giannis Antetokounmpo	2995421	13564383	-10568962	Bradley Beal	22116750	12236068	9880681
Rudy Gobert	2121288	12126095	-10004807	Bismack Biyombo	17000000	6821376	10178623
Will Barton	3533333	13166861	-9633528	Mike Conley	26540100	15863738	10676361
Isaiah Thomas	6587132	15107640	-8520508	Allen Crabbe	18500000	5074226	13425773

Lasso Regularization on All Features



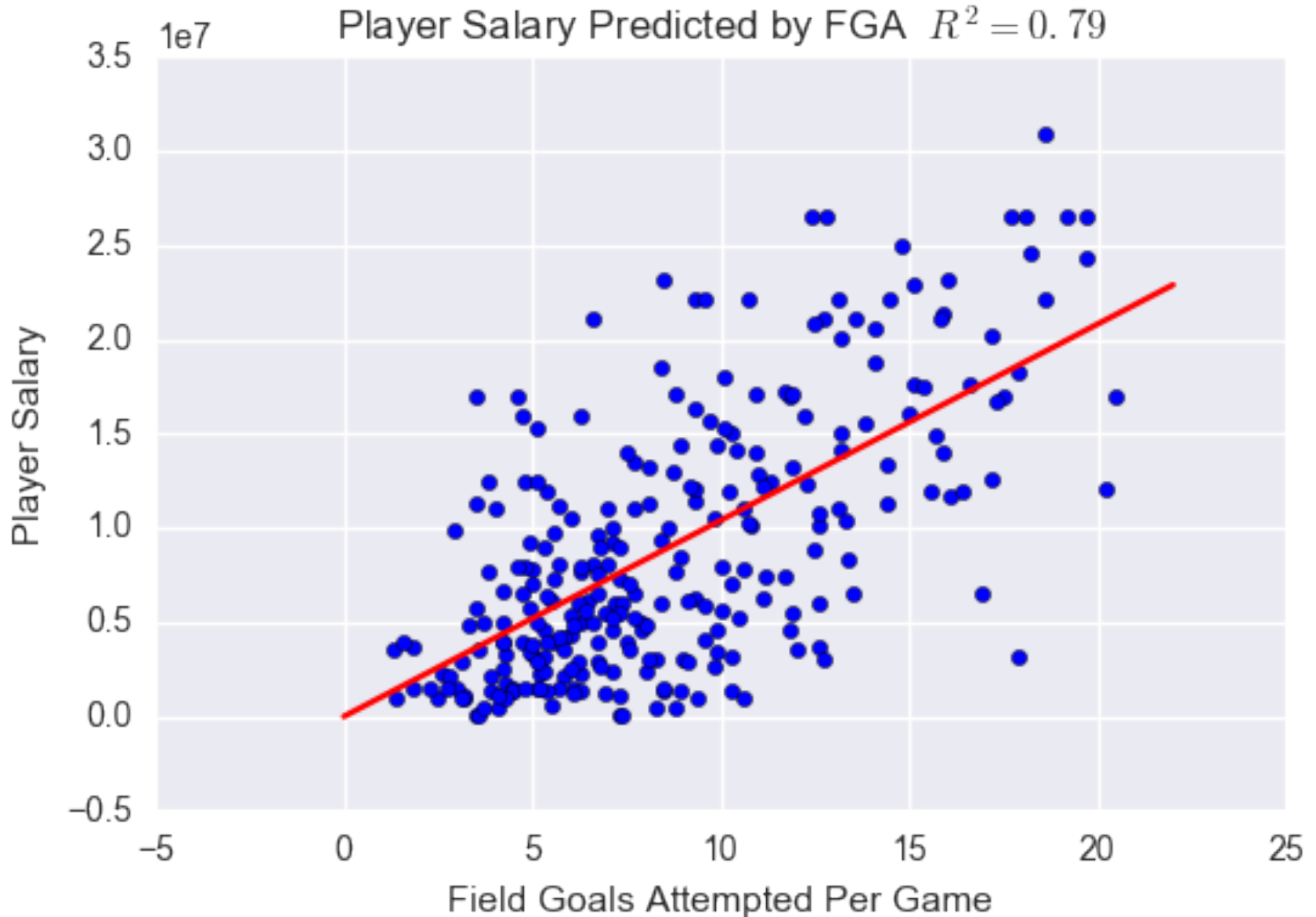
Lasso Model $R^2 = 0.53$



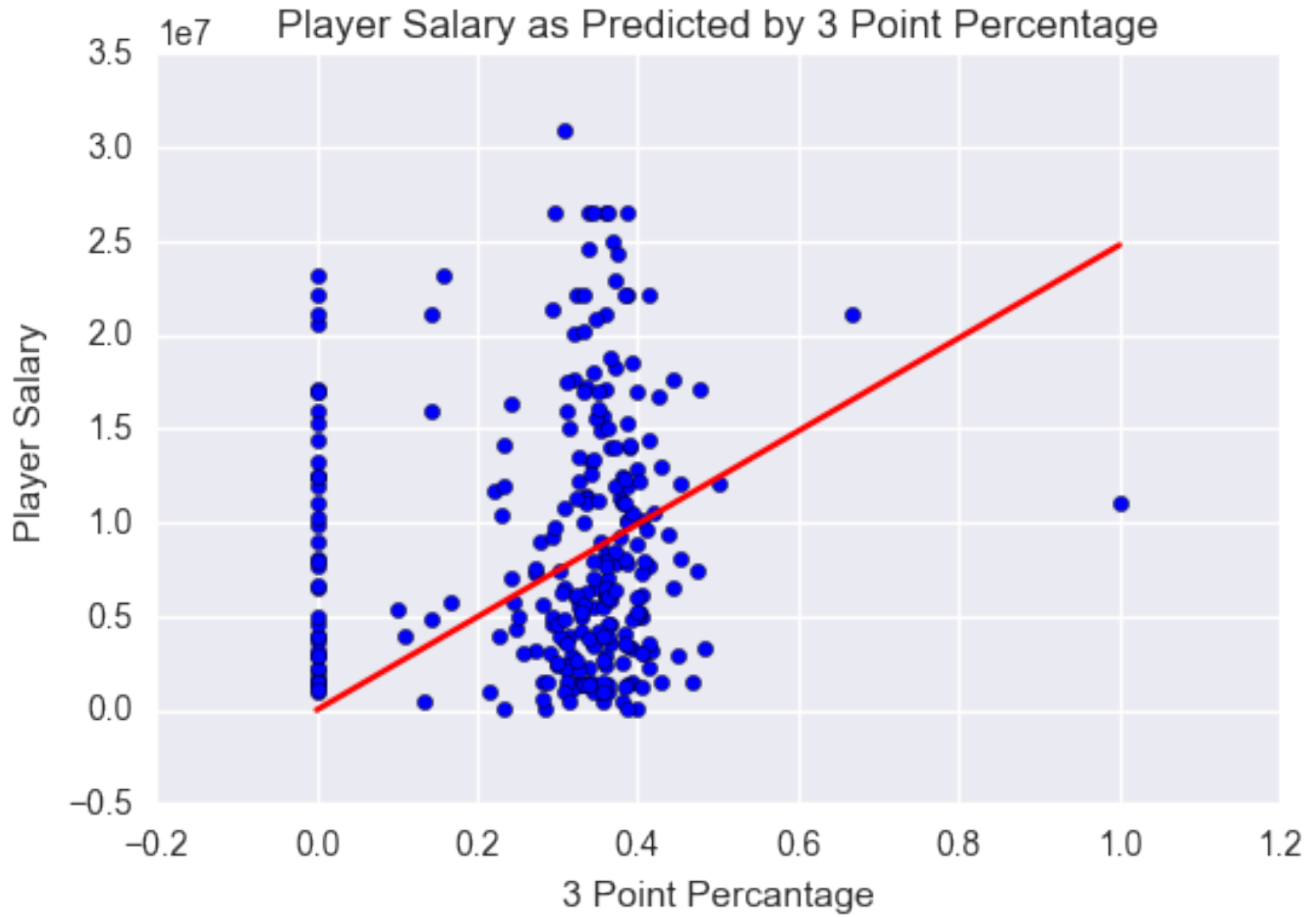
	Coefficient
Feature	
Pos[T.PG]	-1.225892e+07
Tm[T.PHI]	-8.716991e+06
Tm[T.SAC]	-3.555470e+06
Tm[T.LAL]	-3.024502e+06
Tm[T.MIN]	-1.961597e+06
Tm[T.UTA]	-1.922449e+06
Tm[T.TOT]	-1.535185e+06
Tm[T.BOS]	-7.920448e+05
Tm[T.CHI]	7.171023e+05
Tm[T.SAS]	1.346254e+06
Tm[T.LAC]	1.628250e+06
Tm[T.TOR]	2.859012e+06
BLK	2.872529e+06
Tm[T.MIA]	5.462041e+06
FTA	8.767224e+06
GS	9.802988e+06
Tm[T.CLE]	1.128060e+07

Lasso chose mostly Team Features

Players get paid to take shots???



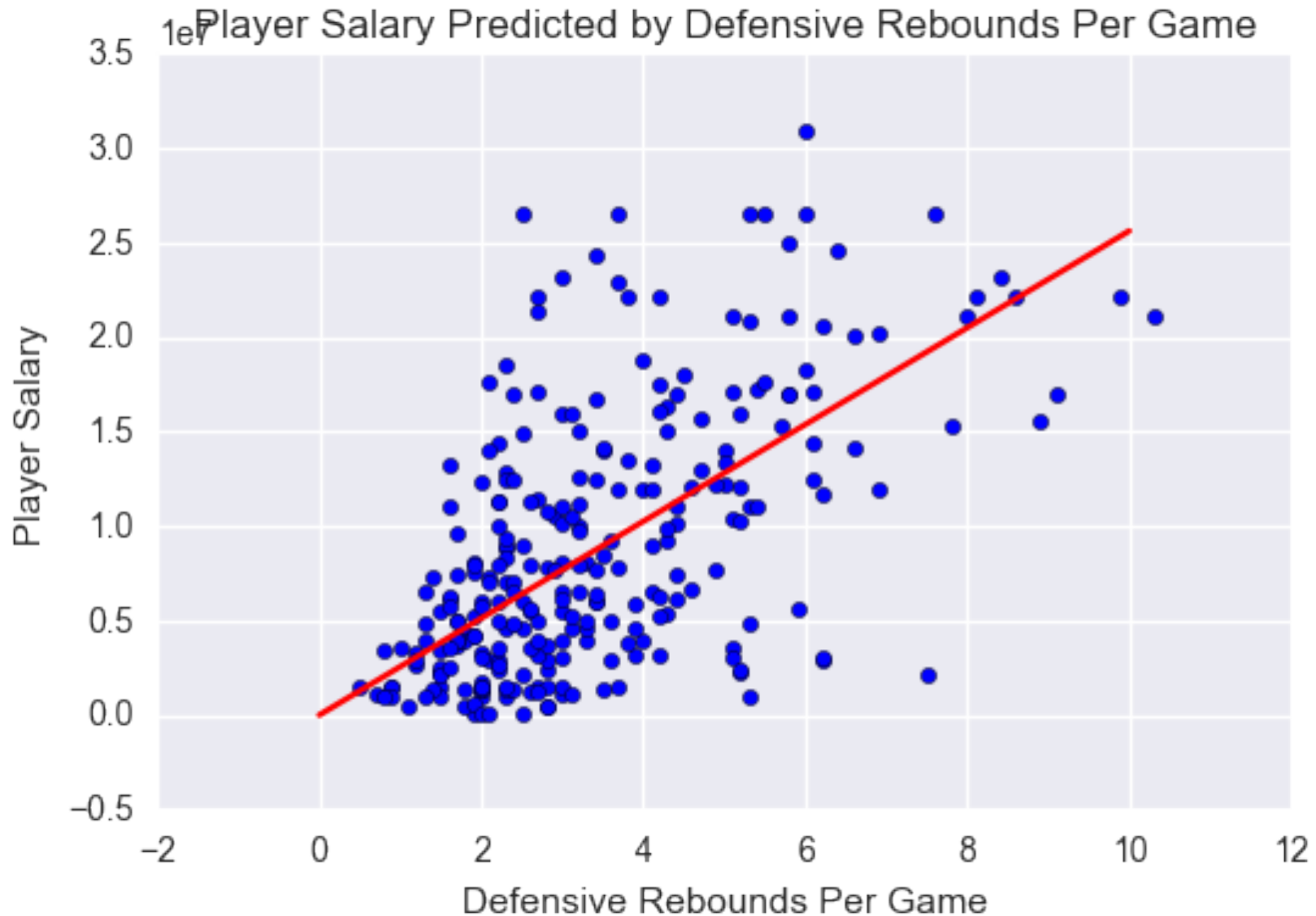
Percentages are meaningless???



What players can improve

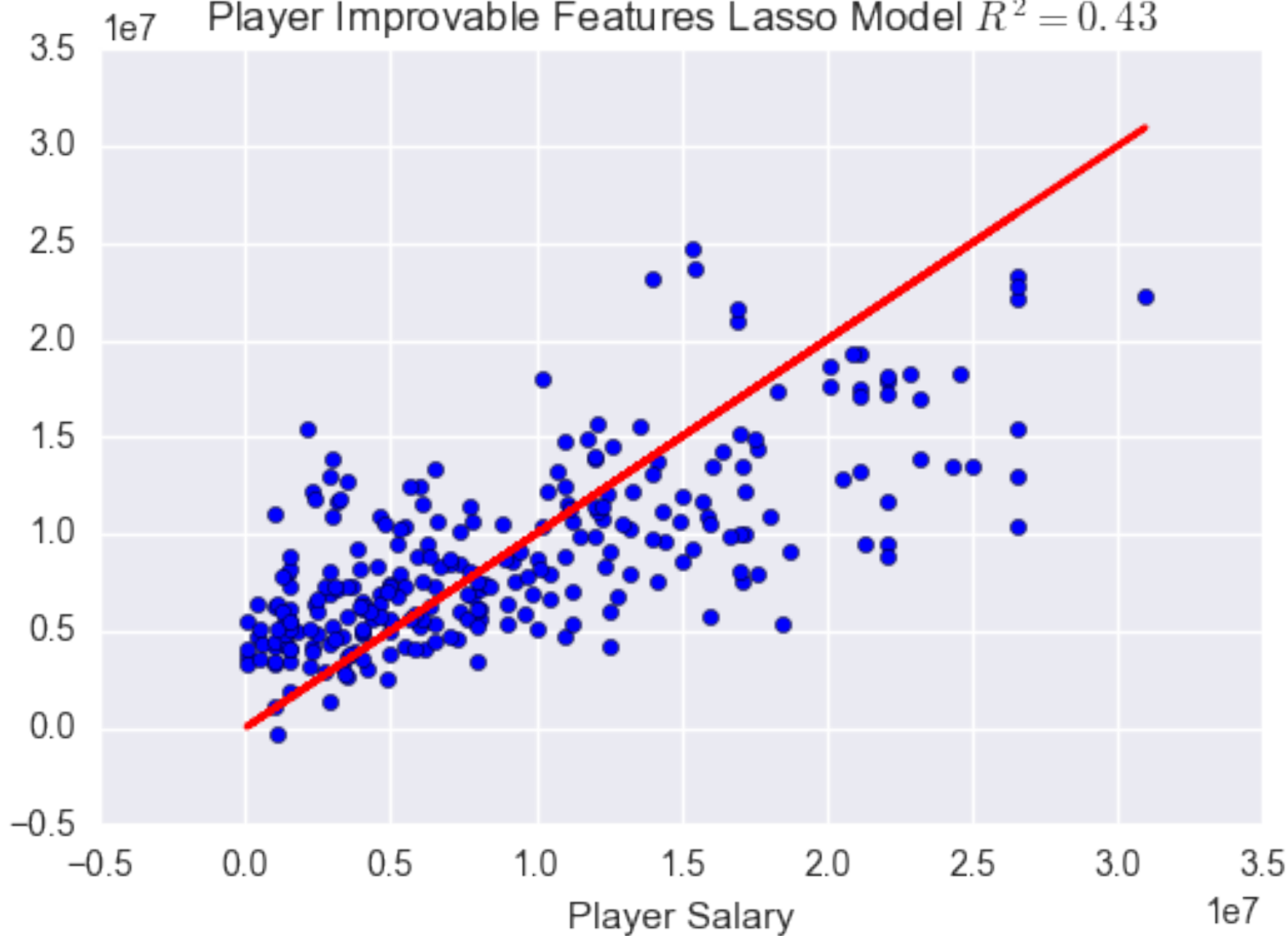
	Correlation
Salary	1.000000
DRB	0.587944
TOV	0.565806
AST	0.443025
STL	0.405477
PF	0.356538
BLK	0.347770
ORB	0.323643
FGP	0.240471
PP2	0.210840
eFGP	0.168928
Weight	0.122842
Height	0.098205
Experience	0.089822
FTP	0.057023
PP3	0.043476
Pos[T.SF]	0.035377
Pos[T.PF-C]	-0.009908
Pos[T.PG]	-0.025253
Pos[T.PF]	-0.041082
Pos[T.SG]	-0.061660

Defensive Rebounds DO Pay



Player Improvable Features Lasso Model $R^2 = 0.43$

Predicted Salary



	Coefficient
Feature	
Pos[T.PG]	-2.449998e+07
PF	-1.840238e+07
Pos[T.PF]	-2.149258e+06
Pos[T.SG]	9.313115e+05
Pos[T.SF]	1.582945e+06
ORB	3.263433e+06
BLK	1.138726e+07
FTP	1.168691e+07
Experience	1.339537e+07
PP3	1.350167e+07
TOV	2.091107e+07
AST	6.842915e+07
DRB	9.549986e+07

Lasso Coefficients Improvable Features

Conclusions

- NBA Player Salaries fairly predictable: $R^2 = 0.53$
- Team Membership, Field Goal Attempts surprisingly predictive
- Meaningful stats also predictive: $R^2 = 0.43$
- Most meaningful stats counter intuitive:
Defensive Rebounds > 3 Pointer Percentage

Thanks!

