Estimating (Four) Factor Values in the NBA: A Seemingly Unrelated Regression Analysis

Jonathan Bosch

Dax Speakman

Dr. Shane Sanders

Background

- [1]Dean Oliver's Four Factors of Basketball Success:

Effective Field Goal Percentage (eFG%)
Turnover Rate (TOV%)
Rebounding Rate (TRB%)
Free Throw Rate (FTr)

- A unit improvement in an offensive factor is equal to a defensive factor unit improvement (in terms of score margin implication)

- [2]Research shows offensive factor improvements are 2.5 times more valuable than defensive factor improvements in the NBA free agent market.

- **Can we find a difference in the NBA free agency market valuation of the factors such that a win-maximizing team can exploit?**

- Can players whose win value arise from relatively expensive factors be shed, in favor of those whose win value arise from relatively inexpensive factors?
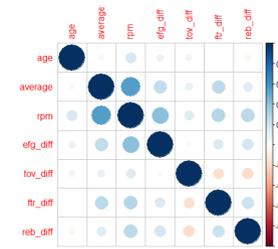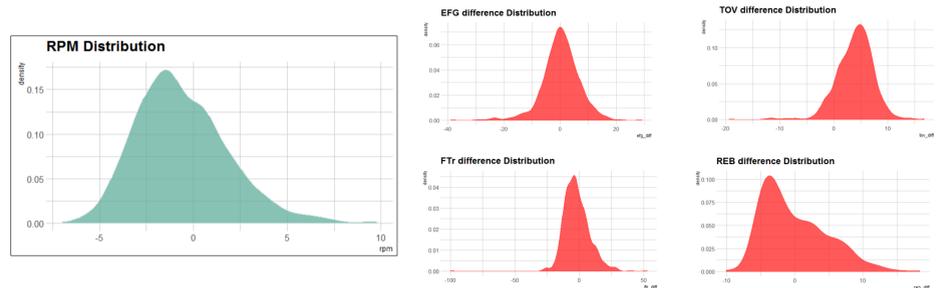
Data

- Contract details[3]
  - Player name, age, position, team (signed to), year (2014-2018), length of contract, yearly average salary

- Player Stats[4] (from Year - 1)
  - Player name, season (2013-2017), games, eFG% TOV% FTr (and for opponent), TRB%

- Real Plus-Minus[5] (from Year - 1)
  - Player name, season (2013-2017), RPM

- Four Factors manipulation
  - eFG difference = eFG% - eFG%$_{opponent}$
  - TOV difference = TOV% - TOV%$_{opponent}$
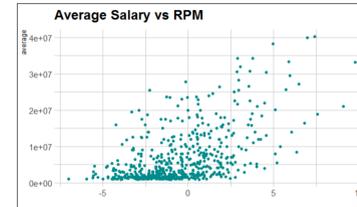 - FTr difference = FTr - FTr$_{opponent}$
  - REB difference = TRB% - 10.0643

*10.0643 is the average total rebounding rate in the NBA

player_season	position	age	to	average	rpm	efg_diff	tov_diff	ftr_diff	reb_diff
ChrisPaul2017	PG	33	HOU	39932648	6.99	2.6387249	6.0	9.2638621	-0.5643
LeBronJames2017	SF	33	LAL	38328212	4.96	4.2434881	1.3	7.9091704	3.0357
PaulGeorge2017	SF	28	OKC	34227984	2.90	-1.0026253	5.7	6.0114701	-1.4643
KevinDurant2017	SF	30	GSW	30750000	3.61	7.5955357	3.0	10.4048544	1.1357

Exploratory Data Analysis

What is the worth of these factor differentials in free agency?

- Sampled data into train (80%) & test (20%) sets

- We want to estimate the following two equations in order to:

  a) Estimate the weights in which our four factor differentials compute to RPM

  b) Estimate the value (in salary) of one unit of RPM

$$RPM = \beta_0 + \beta_1 \cdot eFG_{diff} + \beta_2 \cdot TOV_{diff} + \beta_3 \cdot FTr_{diff} + \beta_4 \cdot REB_{diff}$$

$$Average\ Salary = \beta_0 + \beta_1 \cdot RPM + \beta_2 \cdot AGE + Positions_{factor} + Team_{factor}$$

- Anticipate correlated errors between the two equation
- Chose to run the two regressions simultaneously
- "systemfit" package - run seemingly unrelated regression[6] (SUR) in R

	RPM	Average Salary
RPM	1.0	-0.048
Average Salary	-0.048	1.0

← The correlations of the residuals after SUR analysis

RPM

Variable	Coefficient (β)
Intercept	*-1.140*
eFG$_{diff}$	*0.126*
TOV$_{diff}$	*0.145*
FTr$_{diff}$	*0.049*
REB$_{diff}$	*0.088*

- All variables significant at 99% confidence
- RMSE: 2.1
- R-squared: 0.26

Average Salary

Variable	Coefficient (β)
Intercept	*-13,207,208*
RPM	*1,754,800**
Age	*-192,462**
PG	*-719,087*
SG	*455,403*
SF	*435,743*
PF	*-1,470,012*

- * significant at 95%, ** significant at 99%
- RMSE: 6,217,319
- R-squared: 0.35

We apply chain rule to calculate the following factor differential values:

eFG$_{diff}$ = \$221,104.80   **FTr$_{diff}$ = \$85,985.20**

TOV$_{diff}$ = \$254,446.00   **REB$_{diff}$ = \$154,422.40**

- Normal distributions
- Average (Salary) ~ RPM anticipated correlation not seen as an issue

How much are teams actually paying for factor differentials?

$$Average\ Salary = \beta_0 + \beta_1 \cdot eFG_{diff} + \beta_2 \cdot TOV_{diff} + \beta_3 \cdot FTr_{diff} + \beta_4 \cdot REB_{diff} + \beta_5 \cdot Age + Positions_{factor} + Team_{factor}$$

We compare these coefficients with the factor differential values we previously calculated.

Factor	Market Price	Market Value	Percent Overpaid
eFG$_{diff}$	\$195,000	\$221,000	-11.8 % (underpaid)
TOV$_{diff}$	\$281,000	\$254,000	10.6 %
FTr$_{diff}$	\$182,000	\$86,000	111.6 %
REB$_{diff}$	\$371,000	\$154,000	141.0 %

Variable	Coefficient (β)
Intercept	*4,205,598*
eFG$_{diff}$	*195,351**
TOV$_{diff}$	*280,772**
FTr$_{diff}$	*182,468**
REB$_{diff}$	*371,481**
Age	*-61,817*
PG	*3,369,446*
SG	*2,608,987*
SF	*3,827,043**
PF	*-418,686*

- * significant at 95%, ** significant at 99%
- RMSE: 6,903,649
- R-squared: 0.20

Conclusions

- Holding factor values constant, PF the cheapest on the FA market
  → Construct cheaper, role-players at the PF position

- REB and FT factors are much more overpaid than EFG and TOV
  → EFG and TOV not salient because of opponent factor

- PG strongly overpaid - better estimated by RPM
  → Factors blame PG for TOV, but don't credit for passing/playmaking

Future Work

- Similar analysis with 8 factors – break up the differentials
  → Is Offense or Defense driving these valuations?

- 9 Factors – passing/playmaking component
  → Too important part of the game to ignore at the player level

References

1. Oliver, Dean. Basketball on Paper: Rules and Tools for Performance Analysis. Potomac Books, Inc., 2011.
2. Ehrlich, J., Sanders, S., & Boudreaux, C. J. (2019). The relative wages of offense and defense in the NBA: A setting for win-maximization arbitrage? Journal of Quantitative Analysis in Sports. https://doi.org/10.1515/jqas-2018-0095
3. "NBA Contracts." Spotrac.com, www.spotrac.com/nba/contracts/.
4. "NBA Player Stats." NBA Stats, stats.nba.com/players/.
5. "ESPN RPM." ESPN, ESPN Internet Ventures, 19 June 2019, www.espn.com/nba/statistics/rpm.
6. Perform Seemingly Unrelated Regression in R. UCLA: Statistical Consulting Group. from https://stats.idre.ucla.edu/r/faq/how-can-i-perform-seemingly-unrelated-regression-in-r/

Contact

jbosch@syr.edu   djspeakm@syr.edu   sdsander@syr.edu