A New Framework to Estimate Return on Investment for Player Salaries in the National Basketball Association

Jackson P. Lautier, PhD, FSA, CERA, MAAA[1]

[1]Department of Mathematical Sciences, Bentley University

Southern Methodist University
Department of Statistics and Data Science
January 26, 2024
Update: March 2024

1. Introduction
2. Methods
   - Literature Review
   - WinLogit
   - Single Game Value
   - Return on Investment
3. Results
   - On-court Only
   - ROI Calculations
4. Discussion

Table of Contents

A note on this project

First audience of substantial revisions to the working paper:

Lautier, J. P. (2023). A new framework to estimate return on investment for player salaries in the National Basketball Association. [arXiv] [github]

The following has benefited from two anonymous reviewers at JQAS, participants at the 2023 Midwest Sports Analytics Symposium, and personal correspondence with Seth Partnow.

[Click to download latest version of manuscript]

Crossovers to cash flows

- Standard financial practice relies on measuring realized returns to evaluate performance (e.g., Fund A vs. Fund B)
- Typically, "investment in" and "returns out" are both monetary $\implies$ calculations just Finance 101
- Difficulties arise when the return is non-monetary; e.g., "returns out" = performance on a basketball court
- Goal: A method to translate performance on the basketball court into cash flows (i.e., *crossovers* to cash flows) $\implies$ allows for standard return-on-investment (ROI) calculations
- Recent NBA team valuations hit $4B (Wojnarowski, 2022) $\implies$ such methods taking on greater importance

BENTLEY UNIVERSITY

Converting *buckets* into bucks

Goal: Convert a player's per-game on court performance (i.e., made baskets, rebounds, etc.) into a series of cash flows:

I. Select Time Period

Decide on a measurement time horizon (e.g., one NBA regular season)

→

II. Redistribute Credit

Determine an optimization goal (i.e., wins) and distribute performance-based credit to each player, for each game over the time period in I

→

III. Estimate SGV

Estimate a dollar value of each game in the measurement time horizon (each game need not have the same value)

→

IV. Create Cash Flows

Take an element-wise product of the vectors in II and III to create a series of realized cash flows per player over the measurement time horizon

→

V. Perform ROI Calculations

Treat the player salary as an invested (−) cash flow with IV as the realized (+) cash flows to perform the desired financial analysis

Figure: NBA contractual ROI estimation framework summary

Table of Contents

BENTLEY UNIVERSITY

What has been done and what we need

- Need: single-game unit calculations ("*one game*, the unit of basketball that is most important" (Oliver, 2004)).
- Advantages: standardization (i.e., per possession), able to use logistic regression, salary initial investment $\rightarrow$ bad games ($-$) and missed games (defaults)
- Despite substantial related literature (e.g., Berri, 1999; Page et al., 2007; Fearnhead and Taylor, 2011; Casals and Martínez, 2013; Berri et al., 2007; Berri and Bradbury, 2010; Martínez, 2012; Lackritz and Horowitz, 2021; Idson and Kahane, 2000; Tunaru et al., 2005; Terner and Franks, 2021), no existing approach is (1) per-game, (2) updated with player tracking data, and (3) considers both player salary and performance together.

BENTLEY UNIVERSITY

Two of note

- Game Score (GmSc) (Sports Reference LLC, 2023b):

$$\begin{aligned} \text{GmSc} = {} & \text{PTS} + 0.4\text{FG} - 0.7\text{FGA} - 0.4(\text{FTA} - \text{FT}) \\ & + 0.7\text{ORB} + 0.3\text{DRB} + \text{STL} + 0.7\text{AST} \\ & + 0.7\text{BLK} - 0.4\text{PF} - \text{TOV}, \end{aligned}$$

- Win Score (WSc) (Berri et al., 2007)

$$\begin{aligned} \text{WSc} = {} & \text{PTS} + \text{ORB} + \text{DRB} + \text{STL} + 0.5\text{BLK} \\ & + 0.5\text{AST} - \text{FGA} - 0.5\text{FTA} - \text{TOV} - 0.5\text{PF}, \end{aligned}$$

Both versions of player evaluation metrics above will also be tracked, in addition to the *WinLogit* we propose.

BENTLEY UNIVERSITY

Converting *minutes* into moolah

Goal: Convert a player's per-game on court performance (i.e., made baskets, rebounds, etc.) into a series of cash flows:

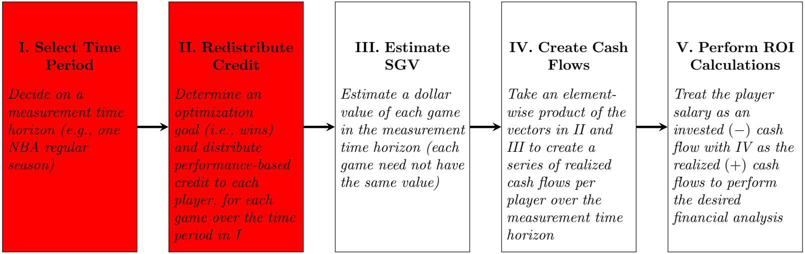

Figure: NBA contractual ROI estimation framework summary

Building the WinLogit

(1) Establish principles to build/calibrate model, select data
(2) Attractiveness of logistic regression for basketball data
(3) Fitting the model
(4) Build a wealth redistribution model from (3)

WinLogit: Establish principles, select data

Principles:

- *Edwardsian in Outcome* ("play to win the game")
- *Value all Activity* (e.g., distance traveled)
- *No Double Counting* (e.g., adjusting passes made)

The starting 36 statistical categories: FG2O, FG2X, FG3O, FG3X, FTMO, FTMX, PF, STL, AORB, ADRB, AST, BLKS, TO, BLKA, PFD, SAST, DEFL, CHGD, AC2P, C3P, OBOX, DBOX, OLBR, DLBR, DFGO, DFGX, DRV, ODIS, DDIS, APM, AST2, FAST, OCRB, DCRB, AORC, and ADRC.

BENTLEY UNIVERSITY

Logistic regression (Kutner et al., 2005)

Let $y_i = 1$ (win) or $y_i = 0$ (loss) with probability $\Pr(y_i = 1 \mid \boldsymbol{x}_i, \boldsymbol{\beta}) \equiv p_i$, where $\boldsymbol{x}_i = (1, X_{i1}, \ldots, X_{ik})$ is a row of the design matrix of team level statistics, $\mathbf{X}$.

That is, $y_i \sim \text{Bern}(p_i)$ for $i = 1, \ldots, n$. The binary logit regression model has the form, for $i = 1, \ldots, n$,

$$f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = \frac{\exp(y_i \boldsymbol{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})},$$

or,

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{x}_i^\top \boldsymbol{\beta}. \tag{1}$$

The form (1) implies

$$p_i = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})} = \frac{1}{1 + \exp(-\boldsymbol{x}_i^\top \boldsymbol{\beta})}.$$

Logistic regression (cont.)

- Regression coefficients are called log-odds ratios
- Stats Interpretation: β_j is the additive increase in the log-odds success probability from a unit increase in x_{ij}, when all other x_{ij^*}'s, $j^* \neq j$, are held fixed, $j, j^* = 1, \ldots, k$.
- Basketball Interpretation: Any field in $\mathbf{X}$ that returns a positive (and significant) β_j has a positive contribution to win probability and vice versa (*Edwardsian in Outcome*)

Logistic regression (cont.)

Logistical regression in the context of a basketball game outcome offers some pleasing interpretations. First, if we center each covariate, X_{ij}, i.e., replace X_{ij} with $(X_{ij} - \bar{X}_j)$, where $\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$, then the intercept, β_0, becomes the logit at the mean. In other words, an average game by a team yields a

$$p(\bar{X}_1, \ldots, \bar{X}_k) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)},$$

probability of winning.

Hence, $\beta_0 = 0$ implies $p(\bar{X}_1, \ldots, \bar{X}_k) = 0.5$, a quite reasonable assumption (more on this soon).

Basketball data, a closer look

For any game,[1] i, $i = 1, \ldots, n$,

Player	F2MO	F2MX	$\ldots$
Tatum	X_{i11}	X_{i21}	$\ldots$
Brown	X_{i12}	X_{i22}	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$
Celtics	$X_{i1\bullet} = \sum_{m=1}^{15} X_{i1m}$	$X_{i2\bullet} = \sum_{m=1}^{15} X_{i2m}$	$\ldots$

Hence, **if** $\beta_0 = 0$,

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i^\top \boldsymbol{\beta} \implies \text{logit}(p_i) = \sum_{m=1}^{15} \text{logit}(p_{im}),$$

where p_i is the win probability for game outcome i, $i = 1, \ldots, n$, and p_{im} is the win probability for player m, $m = 1, \ldots, 15$.

[1] Almost, shot clock violations are team turnovers.

Theorem 1: Team and Player Logits

Let X_{ijm} represent the individual total for player m, $m = 1, \ldots, 15$, for statistical category $j = 1, \ldots, k$ for game outcome i, $i = 1, \ldots, n$. Fix $j = 1, \ldots, k$ and define the team total statistics for game outcome i, $i = 1, \ldots, n$, as

$$X_{ij\bullet} = \sum_{m=1}^{15} X_{ijm}.$$

Then

$$X_{ij\bullet} - \bar{X}_{ij\bullet} = \sum_{m=1}^{15} \left(X_{ijm} - \bar{X}_{ijm} \right),$$

where $\bar{X}_{ij\bullet} = \frac{1}{n} \sum_{i=1}^{n} X_{ij\bullet}$ and $\bar{X}_{ijm} = \frac{1}{15n} \sum_{i=1}^{n} \sum_{m=1}^{15} X_{ijm}$. Further, if we assume $\beta_0 = 0$ and recall (1), then

$$\text{logit}(p_i) = \sum_{m=1}^{15} \text{logit}(p_{im}),$$

where p_i is the win probability for game outcome i, $i = 1, \ldots, n$, and p_{im} is the win probability for player m, $m = 1, \ldots, 15$,

$$p_{im} = \frac{\exp(\boldsymbol{x}_{im}^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_{im}^\top \boldsymbol{\beta})},$$

where $\boldsymbol{x}_{im}^\top = (X_{i1m} - \bar{X}_{i1m}, \ldots, X_{ikm} - \bar{X}_{ikm})^\top$. That is, the team level logit of the win probability may be written as a sum of the logits of the individual player win probabilities.

The fitted model

Field	Coefficient Estimate	Standard Error	Significance	Variable Importance
FG2O	0.251	0.0267	$***$	9.40
FG2X	-0.349	0.0274	$***$	12.73
FG3O	0.537	0.0368	$***$	14.62
FG3X	-0.368	0.0283	$***$	13.01
FTMO	0.122	0.0221	$***$	5.52
FTMX	-0.220	0.0350	$***$	6.31
PF	-0.197	0.0224	$***$	8.76
AORB	0.356	0.0437	$***$	8.15
ADRB	0.316	0.0246	$***$	12.84
STL	0.443	0.0354	$***$	12.52
BLK	0.132	0.0336	$***$	3.92
TOV	-0.347	0.0292	$***$	11.85
PFD	0.214	0.0329	$***$	6.51
SAST	0.076	0.0214	$***$	3.56
CHGD	0.522	0.1008	$***$	5.18
AC2P	0.041	0.0117	$***$	3.48
C3P	-0.067	0.0140	$***$	4.81
DBOX	0.053	0.0242	$*$	2.18
DFGO	-0.230	0.0179	$***$	12.81
DFGX	0.086	0.0133	$***$	6.50
DDIS	-1.000	0.2009	$***$	4.98
APM	0.016	0.0031	$***$	5.25
OCRB	0.290	0.0371	$***$	7.81
DCRB	0.338	0.0338	$***$	9.99

Table: **WinLogit Logistic Regression Model Parameters**. Based on team outcomes for the 2022-2023 NBA regular season ($n = 2{,}452$). Significant at $\alpha = 0.001$ ($***$), $\alpha = 0.01$ ($**$), and $\alpha = 0.05$ ($*$). The McFadden R^2 (McFadden, 1974) is 0.6457. Variable importance computed using Kuhn (2008).

Fitted model notes

- We find $\hat{\beta}_0 = -0.004930$ with a p-value of 0.948 $\implies$ forcing $\beta_0 = 0$ results in minimal bias (i.e., total estimated win probability is 1226.88, unbiased is 1226)
- Coefficient significance, size and direction plausible
- Centering player level data: $\text{logit}(p_{im}) = 0$ for an average game, $\text{logit}(p_{im}) < 0$ for a below average game, and $\text{logit}(p_{im}) > 0$ for an above average game $\implies$ pleasing *replacement player* interpretation
- Logistic regression framework: both teams in a single game compete to obtain the largest team logit, with individual players making both positive and negative contributions

Defining the WinLogit

- Restrict ROI calculations to on court performance only; i.e., set of players with playing time
- Assume a game is worth 1 unit
- All players with playing time (i.e., `minutes` > 0), denoted $\mathcal{M}_g$ for game g, $g = 1, \ldots, n/2$, begin with a $1/\bar{m}$ share, where $\bar{m} = m^*/n$ and m^* is the total number of players with playing time in the n total games (i.e., $m^* = \sum_g \sum_{m \in \mathcal{M}_g} 1$)
- Interpretation: average game results in a $1/\bar{m}$ share
- Comparison: standardize the measure

WinLogit Definition

Define the basic sample statistics

$$\overline{\mathrm{WL}}_{m^*}(\boldsymbol{\beta}) = \frac{1}{m^*} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \mathrm{logit}(p_{gm}),$$

and

$$s(\mathrm{WL})_{m^*}(\boldsymbol{\beta}) = \sqrt{\frac{1}{m^* - 1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(\mathrm{logit}(p_{gm}) - \overline{\mathrm{WL}}_{m^*} \right)^2}.$$

Then, we define the *WinLogit* for player $g \in \mathcal{M}_g$ in game g, $g = 1, \ldots, n/2$, denoted $\mathrm{WinLogit}_{gm}$, as

$$\mathrm{WinLogit}_{gm}(\boldsymbol{\beta}) = \frac{1}{s(\mathrm{WL})_{m^*}} \left(\mathrm{logit}(p_{gm}) - \overline{\mathrm{WL}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}. \quad (2)$$

Theorem 2: WinLogit Properties

Let the WinLogit_{gm} take the form of (2) for player $m \in \mathcal{M}_g$, $g = 1, \ldots, n/2$. Then the WinLogit_{gm} is standardized such that

$$\frac{1}{m^*} \sum_{i=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{WinLogit}_{gm}$$
$$= \sqrt{\frac{1}{m^* - 1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(\text{WinLogit}_{gm} - \frac{1}{\bar{m}} \right)^2} = \frac{1}{\bar{m}}.$$

Further, let $\hat{\beta}_{\text{MLE}}$ be the MLE of the logistic regression assumed in Theorem 1. Then the MLE of $\text{WinLogit}_{gm}(\beta)$ is $\text{WinLogit}_{gm}(\hat{\beta}_{\text{MLE}})$.

Wealth redistribution interpretation

In an economic interpretation, the WinLogit may be thought of as a type of wealth redistribution tool.

Starting from the assumption all players in a game have an average performance and thus a perfect uniformity of wealth, the WinLogit then redistributes the wealth to each player based on each player's on court performance in comparison to an average (or replacement) player.

Note, there is a bias correction as

$$\sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{WinLogit}_{gm} = \frac{n}{2},$$

by definition.

For comparison

Similarly, we may define for player $m \in \mathcal{M}_i$ in game i, $i = 1, \ldots, n$

$$\text{GmSc}^*_{gm} = \frac{1}{s(\text{GS})_{m^*}} \left(\text{GmSc}_{gm} - \overline{\text{GS}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}},$$

and

$$\text{WnSc}^*_{gm} = \frac{1}{s(\text{WS})_{m^*}} \left(\text{WnSc}_{gm} - \overline{\text{WS}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}},$$

where $\overline{\text{GS}}_{m^*}$, $\overline{\text{WS}}_{m^*}$, $s(\text{GS})_{m^*}$, and $s(\text{WS})_{m^*}$ take the usual sample statistic forms.

WinLogit robustness analysis

Rank	Team	Wins	WL (ae)	WS (ae)	GS (ae)	WLR (ae)	WSR (ae)	GSR (ae)
	Median Error		3.66	4.95	4.82	1.00	3.00	4.00
	Average Error		5.49	5.99	6.47	2.87	3.93	4.87
1	MIL	58	46.08 (11.9)	45.08 (12.9)	42.13 (15.9)	1 (0)	2 (1)	9 (8)
2	BOS	57	45.78 (11.2)	45.60 (11.4)	43.71 (13.3)	2 (0)	1 (1)	2 (0)
⋮								
14	LAL	43	41.96 (1.0)	42.74 (0.3)	42.22 (0.8)	12 (2)	8 (6)	8 (6)
15	NOP	42	41.56 (0.4)	41.27 (0.7)	41.40 (0.6)	15 (0)	14 (1)	14 (1)
16	ATL	41	41.24 (0.2)	42.69 (1.7)	43.10 (2.1)	17 (1)	9 (7)	4 (12)
17	MIN	41	40.26 (0.7)	40.00 (1.0)	40.54 (0.5)	21 (4)	22 (5)	20 (3)
⋮								
29	SAS	21	33.67 (12.7)	35.96 (15.0)	37.05 (16.1)	29 (0)	29 (0)	29 (0)
30	DET	17	32.68 (15.7)	34.37 (17.4)	36.18 (19.2)	30 (0)	30 (0)	30 (0)

Table: **Model Versus Actual Wins**. A comparison of actual vs. estimated wins using the winLogit (WL), the Game Score (GS), and the Win Score (WS) models. Results for the 2022-2023 NBA regular season.

WinLogit robustness analysis (cont.)

Field	Coefficient	Standard Error	Test Statistic	Significance
(Intercept)	-14.278	0.6328	-22.56	$***$
WinLogit	17.811	1.1961	14.89	$***$
WnSc*	10.502	2.5387	4.14	$***$
GmSc*	0.884	2.2568	0.39	

Table: **Team Level Models and Wins**. Data for the 2022-2023 NBA regular season. Significant at $\alpha = 0.001$ $(***)$, $\alpha = 0.01$ $(**)$, $\alpha = 0.05$ $(*)$, and $\alpha = 0.10$ $(\cdot)$. The McFadden R^2 (McFadden, 1974) is 0.5203.

Comparing distributions

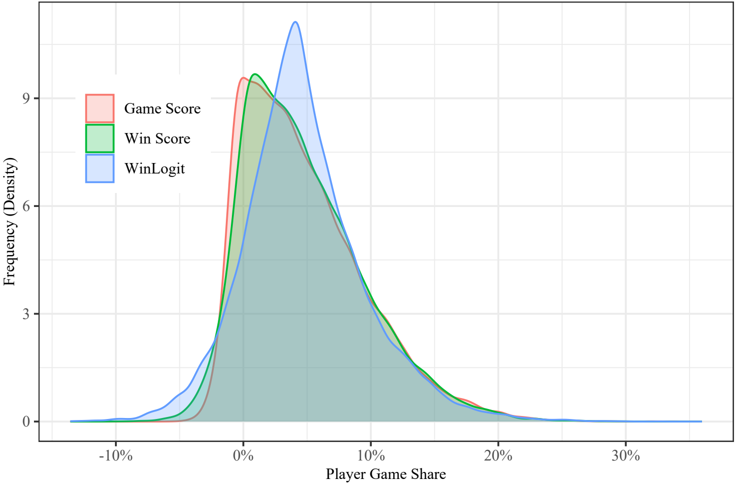

Figure: Wealth Redistribution Comparison (22-23 NBA Regular Season)

Converting *dunks* into dough

Goal: Convert a player's per-game on court performance (i.e., made baskets, rebounds, etc.) into a series of cash flows:

I. Select Time Period
Decide on a measurement time horizon (e.g., one NBA regular season)

→

II. Redistribute Credit
Determine an optimization goal (i.e., wins) and distribute performance-based credit to each player, for each game over the time period in I

→

III. Estimate SGV
Estimate a dollar value of each game in the measurement time horizon (each game need not have the same value)

→

IV. Create Cash Flows
Take an element-wise product of the vectors in II and III to create a series of realized cash flows per player over the measurement time horizon

→

V. Perform ROI Calculations
Treat the player salary as an invested (−) cash flow with IV as the realized (+) cash flows to perform the desired financial analysis

Figure: NBA contractual ROI estimation framework summary

What is an NBA game worth?

For $g = 1, \ldots, n/2$, let

$$\begin{aligned}\mathsf{SGV}_g(\boldsymbol{\alpha}) &= \alpha_1 \mathsf{GATE}_g + \alpha_2 \mathbf{1}_{\mathsf{ESPN}} + \alpha_3 \mathbf{1}_{\mathsf{TNT}} \\ &\quad + \alpha_4(\mathbf{1}_{\mathsf{ESPN}} + \mathbf{1}_{\mathsf{TNT}} + \mathbf{1}_{\mathsf{NBATV}}),\end{aligned}$$

where the parameter vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$ consists of α_1, the average ticket price for an NBA regular season game, α_2, the average TV contract revenue for a regular season NBA game on ESPN, α_3, the average TV contract revenue for a regular season game on TNT, and, α_4, the average advertising revenue for a televised regular season game. Further, GATE_g is a random variable that represents the attendance for game g, and $\mathbf{1}_q$ is an indicator function that equals 1 if statement q is true and 0 otherwise.

What is an NBA game worth? (cont.)

Coefficient	Description	Estimate
α_1	Ticket Price	\$102.64
α_2	ESPN TV Revenue	\$13,861,386
α_3	TNT TV Revenue	\$18,461,538
α_4	Advertising Revenue	\$6,080,586

Table: **Component Estimates of SGV$_g$.** Coefficient estimates of SGV$_g(\boldsymbol{\alpha})$ based on available data for the 2022-2023 NBA regular season (National Basketball Association, 2023; Statista, 2023a,c; Lewis, 2023; Statista, 2023b).

Note: Top 5 teams are LAL (\$908.3M), GSW (\$885.4M), BOS (\$831.1M), PHX (\$766.3M), and PHI (\$708.5M).

BENTLEY UNIVERSITY

Converting *splashes* into scratch

Goal: Convert a player's per-game on court performance (i.e., made baskets, rebounds, etc.) into a series of cash flows:

I. Select Time Period
Decide on a measurement time horizon (e.g., one NBA regular season)

→

II. Redistribute Credit
Determine an optimization goal (i.e., wins) and distribute performance-based credit to each player, for each game over the time period in I

→

III. Estimate SGV
Estimate a dollar value of each game in the measurement time horizon (each game need not have the same value)

→

IV. Create Cash Flows
Take an element-wise product of the vectors in II and III to create a series of realized cash flows per player over the measurement time horizon

→

V. Perform ROI Calculations
Treat the player salary as an invested (−) cash flow with IV as the realized (+) cash flows to perform the desired financial analysis

Figure: NBA contractual ROI estimation framework summary

Finance 101

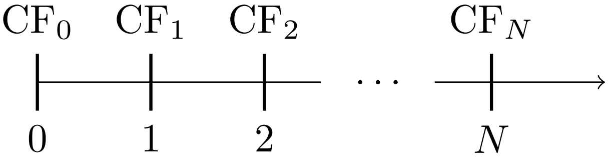

Figure: **Cash flow time line**.

Let CF_0 be the initial investment, and the future (possibly negative) cash flows be $\mathrm{CF}_1, \ldots, \mathrm{CF}_N$. Then the return on investment is the rate, r, such that

$$\mathrm{CF}_0 = \sum_{i=1}^{N} \frac{\mathrm{CF}_i}{(1+r)^i}.$$

NBA ROI (cash flows)

Formally, for any distinct player $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$, let $\mathbf{SGV}_{g \in \mathcal{G}_m} = (\mathrm{SGV}_1, \ldots, \mathrm{SGV}_N)^\top$ be a vector of SGVs, for all games in which player m's team appeared over the investment time horizon, where $\#\{\mathcal{G}_m\} = N \in \mathbb{N}$. Similarly, for the same distinct player $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$, let $\mathbf{WL}_{g \in \mathcal{G}_m} = (\mathrm{WinLogit}^*_{1m}, \ldots, \mathrm{WinLogit}^*_{Nm})^\top$ be a vector of WinLogits for all games in which player m's team appeared over the investment time horizon. Then the vector of return cash flows over the investment time horizon for distinct player $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$ becomes

$$\begin{aligned}\mathbf{CF}_m &= (\mathbf{SGV}_{g \in \mathcal{G}_m})^\top \mathrm{diag}(\mathbf{WL}_{g \in \mathcal{G}_m}) \\ &= (\mathrm{SGV}_1 \mathrm{WinLogit}^*_{1m}, \ldots, \mathrm{SGV}_N \mathrm{WinLogit}^*_{Nm})^\top,\end{aligned}$$

where $\mathrm{diag}(\mathbf{WL}_{g \in \mathcal{G}_m})$ represents a diagonal $N \times N$ matrix with diagonal $\mathbf{WL}_{g \in \mathcal{G}_m}$.

Conditional Unbiasedness

Theorem 3: Conditional Unbiasedness

Let SGV_g be a single game value random variable for any game, $g = 1, \ldots, n/2$ such that $\mathbf{E}(\text{SGV}_g) = \mu \in \mathbb{R}$ for all $g = 1, \ldots, n/2$. Then, conditional on WinLogit_{gm} for all $m \in \mathcal{M}_g$, $g = 1, \ldots, n/2$,

$$\mathbf{E}\left(\sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{SGV}_g \text{WinLogit}^*_{gm} \middle| \text{WinLogit}^*_{gm} \right) = \mu \frac{n}{2}.$$

That is, the wealth redistribution model of the WinLogit, when viewed over all players and games in the investment time horizon, is unbiased to the expected total generated revenue.

NBA ROI (final formula)

Finally, to calculate the ROI, let $\boldsymbol{\nu}_m = ((1+r_m)^{-1}, \ldots, (1+r_m)^{-N})^\top$ be a vector of discount factors at the rate, r_m, where $m \in \{\mathcal{M}_g\}_{1 \le g \le n/2}$ is distinct. Then the contractual ROI for distinct player $m \in \{\mathcal{M}_g\}_{1 \le g \le n/2}$ over the investment time horizon, is the rate, r_m, such that

$$\mathrm{CF}_0^m = (\mathbf{SGV}_{g \in \mathcal{G}_m})^\top \mathrm{diag}(\mathbf{WL}_{g \in \mathcal{G}_m}) \boldsymbol{\nu}_m = \sum_{t=1}^{N} \frac{\mathrm{SGV}_t \mathrm{WinLogit}^*_{tm}}{(1+r_m)^t},$$

where CF_0^m is distinct player m's full salary over the investment time horizon.

Comments

- Assumes player salaries paid in a single lump sum at time zero (adjustments straightforward).
- Does not include playoff games.
- Implicitly weights earlier games more (adjustments less straightforward).
- Straightforward to swap out the WinLogit with Game Score or per game version of Win Score.

Table of Contents

NBA 2022-2023 regular season data via publicly available sources:

(1) Statistics page of the NBA (National Basketball Association, 2023) → custom query wrapper for the `python` package `nba_api` (Patel, 2018).[2]

(2) Player salaries via HoopsHype (2023).

(3) NBA Revenue, attendance, etc. (National Basketball Association, 2023; Statista, 2023a,c; Lewis, 2023; Statista, 2023b).

(4) Player positions via RealGM, L.L.C. (2023); Sports Reference LLC (2023a).

To obtain the data and replication code, please navigate to the public `github` repository at https://github.com/jackson-lautier/nba_roi.

[2] Player tracking data missing for 4 of 1,230 regular season games.

On-court performance only

Define for any $g \in \mathcal{G}_m$,

$$\text{WinLogit}^*_{gm} = \begin{cases} \text{WinLogit}_{gm}, & m \in \mathcal{M}_g \\ 0, & m \notin \mathcal{M}_g. \end{cases} \tag{3}$$

In financial parlance, the form of (3) implies a missed game is a *default* (trivially still unbiased). The season total of (3) for player m is then

$$\text{PVWL}_m = \sum_{g \in \mathcal{G}_m} \text{WinLogit}^*_{gm}. \tag{4}$$

We may consider (4) as a present value of a series of cash flows taking the value of (3) discounted at a zero interest rate.

- The per-game approach helps assess the cost of missed games.
- Positional adjustments to ease comparisons across positions (i.e., the *big man bias*)

The per-game approach & missed games

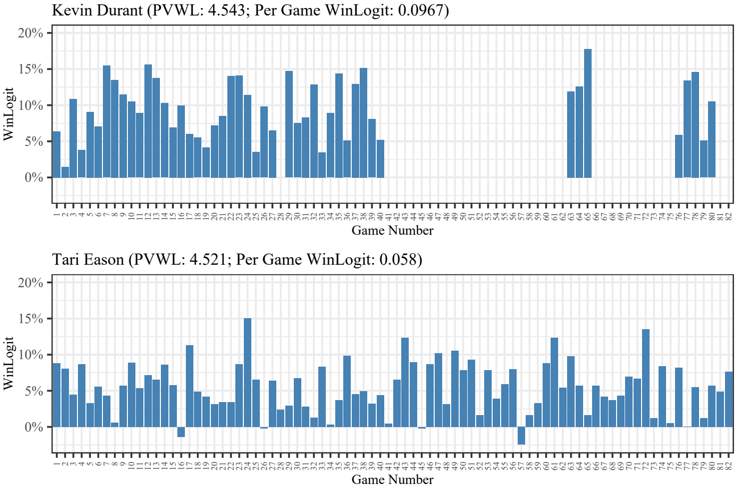

Figure: Quantifying Missed Games: Durant (47 GP) vs. Eason (82 GP)

Top performers (relative to position)

Rank	Player	Position	GP	PVWL
1	Luka Dončić	PG	66	6.159
2	Jayson Tatum	SF	74	6.707
3	Giannis Antetokounmpo	PF	63	8.096
4	Shai Gilgeous-Alexander	PG	68	5.036
5	Nikola Jokić	C	69	10.088
Rank	Player	Position	GP	PVWS
1	Jayson Tatum	SF	74	8.238
2	Luka Dončić	PG	66	8.025
3	Nikola Jokić	C	69	11.505
4	Domantas Sabonis	C	79	11.016
5	Giannis Antetokounmpo	PF	63	7.905
Rank	Player	Position	GP	PVGS
1	Jayson Tatum	SF	74	9.785
2	Nikola Jokić	C	69	10.426
3	Joel Embiid	C	66	10.331
4	Donovan Mitchell	SG	68	7.990
5	Giannis Antetokounmpo	PF	63	8.872

Table: Nikola Jokić is the top overall performer for all three measures. An average player playing all 82 games generates a PV($\cdot$) of 3.896. For complete results, navigate to https://github.com/jackson-lautier/nba_roi.

Biggest disagreements

Name	WL(%)	WS(%)	Name	WL(%)	GS(%)	Name	WS(%)	GS(%)
CJ McCollum	0.31	0.82	Dillon Brooks	0.00	0.72	Jordan Poole	0.66	0.91
Anfernee Simons	0.16	0.65	Anfernee Simons	0.16	0.85	Jaden Ivey	0.55	0.80
Terry Rozier	0.20	0.69	Terry Rozier	0.20	0.87	Jalen Green	0.68	0.92
Dillon Brooks	0.00	0.48	Jaden Ivey	0.14	0.80	Dillon Brooks	0.48	0.72
Killian Hayes	0.12	0.54	Jalen Green	0.28	0.92	Isaiah Hartenstein	0.87	0.65
Jaden Ivey	0.14	0.55	CJ McCollum	0.31	0.94	Andre Drummond	0.79	0.57
Jordan Clarkson	0.21	0.62	Jordan Clarkson	0.21	0.83	Jordan Clarkson	0.62	0.83
Jalen Green	0.28	0.68	Killian Hayes	0.12	0.72	Steven Adams	0.83	0.63
LaMelo Ball	0.22	0.62	RJ Barrett	0.28	0.84	Usman Garuba	0.65	0.45
Fred VanVleet	0.47	0.86	LaMelo Ball	0.22	0.76	Anfernee Simons	0.65	0.85

Table: **Player performance disagreements**. The top ten largest disagreements between PVWL, PVWS, and PVGS for the 2022-2023 NBA regular season in terms of percentile rank (%).

Converting *dimes* into dimes

Goal: Convert a player's per-game on court performance (i.e., made baskets, rebounds, etc.) into a series of cash flows:

I. Select Time Period
Decide on a measurement time horizon (e.g., one NBA regular season)

→

II. Redistribute Credit
Determine an optimization goal (i.e., wins) and distribute performance-based credit to each player, for each game over the time period in I

→

III. Estimate SGV
Estimate a dollar value of each game in the measurement time horizon (each game need not have the same value)

→

IV. Create Cash Flows
Take an element-wise product of the vectors in II and III to create a series of realized cash flows per player over the measurement time horizon

→

V. Perform ROI Calculations
Treat the player salary as an invested (−) cash flow with IV as the realized (+) cash flows to perform the desired financial analysis

Figure: NBA contractual ROI estimation framework summary

Top ROI performers (relative to position, min 42 GP)

			WinLogit	
Rank	Player	Position	SAL	ROI
1	Santi Aldama	PF	$2.09M	53.21%
2	John Konchar	SF	$2.30M	41.78%
3	Jock Landale	C	$1.56M	32.96%
4	Austin Reaves	SG	$1.56M	33.24%
5	Jose Alvarado	PG	$1.56M	16.43%
			Win Score	
Rank	Player	Position	SAL	ROI
1	Santi Aldama	PF	$2.09M	51.98%
2	John Konchar	SF	$2.30M	36.08%
3	Jose Alvarado	PG	$1.56M	16.96%
4	Jock Landale	C	$1.56M	22.46%
5	Austin Reaves	SG	$1.56M	19.90%
			Game Score	
Rank	Player	Position	SAL	ROI
1	Santi Aldama	PF	$2.09M	29.12%
2	Jock Landale	C	$1.56M	25.70%
3	Tyrese Maxey	SG	$2.73M	32.19%
4	Jose Alvarado	PG	$1.56M	19.70%
5	Naji Marshall	SF	$1.78M	20.34%

Table: Santi Aldama is the top overall ROI for the WinLogit and Win Score. Tyrese Maxey is the top overall ROI for Game Score. An average player playing all 82 games with an average salary generates an ROI of 2.71%. For complete results, navigate to https://github.com/jackson-lautier/nba_roi.

BENTLEY UNIVERSITY

Positions as asset classes & risk vs. return

Position	Coefficient of Variation WinLogit*	WnSc*	GmSc*
Center (C)	1.237	1.260	1.905
Power Forward (PF)	1.990	2.327	2.319
Small Forward (SF)	2.070	1.937	1.757
Shooting Guard (SG)	2.176	2.102	2.007
Point Guard (PG)	3.722	2.447	1.990

Table: **Coefficient of Variation for ROI by Position**. A ratio of sample standard deviation to sample mean of 2022-2023 NBA regular season empirical ROI estimates by position.

Table of Contents

BENTLEY UNIVERSITY

Conclusion

- We present the first known attempt to measure an NBA player's contractual return on investment.
- Our results may be used by NBA player personnel decision makers, NBA player agents, and NBA awards voters.
- Forecasting is possible: first project on court performance, then feed the projections into our ROI framework.
- Customization is possible: e.g., swap player credit models, more precise SGVs, more precise player salary CFs, actual game dates, include playoff games, include off court revenue, etc.
- Consider applying to other sports.

Converting *bank shots* into bankrolls

Goal: Convert a player's per-game on court performance (i.e., made baskets, rebounds, etc.) into a series of cash flows:

I. Select Time Period	→	II. Redistribute Credit	→	III. Estimate SGV	→	IV. Create Cash Flows	→	V. Perform ROI Calculations
Decide on a measurement time horizon (e.g., one NBA regular season)		*Determine an optimization goal (i.e., wins) and distribute performance-based credit to each player, for each game over the time period in I*		*Estimate a dollar value of each game in the measurement time horizon (each game need not have the same value)*		*Take an element-wise product of the vectors in II and III to create a series of realized cash flows per player over the measurement time horizon*		*Treat the player salary as an invested (−) cash flow with IV as the realized (+) cash flows to perform the desired financial analysis*

Figure: NBA contractual ROI estimation framework summary

Thank you!

Lautier, J. P. (2023). A new framework to estimate return on investment for player salaries in the National Basketball Association. [arXiv] [github] [Latest version of manuscript]

Jackson P. Lautier, PhD, FSA, CERA, MAAA
e: jlautier@bentley.edu
w: www.jacksonlautier.com

D. J. Berri (1999). "Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association." *Managerial and Decision Economics* **20**, 411–427.

D. J. Berri and J. C. Bradbury (2010). "Working in the land of the metricians." *Journal of Sports Economics* **11**, 29–47.

D. J. Berri, M. B. Short and S. L. Brook (2007). *The Wages of Wins: Taking Measure of the Many Myths in Modern Sport*. Standford Business Books.

M. Casals and J. A. Martínez (2013). "Modelling player performance in basketball through mixed models." *International Journal of Performance Analysis in Sport* **13**, 64–82.

P. Fearnhead and B. M. Taylor (2011). "On estimating the ability of NBA players." *Journal of Quantitative Analysis in Sports* **7**.

HoopsHype (2023). "2022/23 NBA Player Salaries." hoopshype.com. https://hoopshype.com/salaries/players/2022-2023/. Online; accessed 12 June 2023.

T. Idson and L. Kahane (2000). "Team effects on compensation: An application to salary determination in the National Hockey League." *Economic Inquiry* **38**, 345–357.

M. Kuhn (2008). "Building predictive models in R using the caret package." *Journal of Statistical Software* **28**, 1–26.

M. H. Kutner, C. J. Nachtsheim, J. Neter and W. Li (2005). *Applied Linear Statistical Models*. McGraw-Hill Irwin.

J. Lackritz and I. Horowitz (2021). "The value of statistics contributing to scoring in the NBA: A quantitative approach." *The American Economist* **66**, 175–189.

J. Lewis (2023). "NBA announces 9-year extension with ESPN, Turner, through 2025." https://www.sportsmediawatch.com/ - Sports Media Watch https://www.sportsmediawatch.com/2014/10/nba-tv-deal-espn-abc-tnt-nine-year-deal-2025-24-billion-lockout/. Online; accessed 27 December 2023.

J. A. Martínez (2012). "Factors determining production (FDP) in basketball." *Economics and Business Letters* **1**, 21–29.

D. McFadden (1974). "Conditional logit analysis of qualitative choice behavior." *Frontiers in Econometrics* pp. 105–142.

National Basketball Association (2023). "Statistics." nba.com. https://www.nba.com/stats. Online; accessed 13 June 2023.

D. Oliver (2004). *Basketball on Paper: Rules and Tools for Performance Analysis*. Potomac Books, Inc.

BENTLEY UNIVERSITY

G. L. Page, G. W. Fellingham and C. S. Reese (2007). "Using box-scores to determine a position's contribution to winning basketball games." *Journal of Quantitative Analysis in Sports* **3**.

S. Patel (2018). "nba_api: An API client package to access the APIs of NBA.com." GitHub repository. https://github.com/swar/nba_api. Online; accessed 21 July 2023 (v1.2.0).

RealGM, L.L.C. (2023). "2022-2023 NBA Players." RealGM.com. https://basketball.realgm.com/nba/players/2023. Online; accessed 8 February 2024.

Sports Reference LLC (2023a). "2022-23 NBA Player Stats: Totals." Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/leagues/NBA_2023_totals.html. Online; accessed 8 February 2024.

Sports Reference LLC (2023b). "Glossary." Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/about/glossary.html. Online; accessed 26 January 2023.

Statista (2023a). "Gate receipts as percentage of total revenue in the National Basketball Association from 2010/11 to 2022/23." https://www.statista.com/ - Statista https://www.statista.com/statistics/193410/percentage-of-ticketing-revenue-in-the-nba-since-2006/. Online; accessed 27 December 2023.

Statista (2023b). "National Basketball Association sponsorship revenue from 2010 to 2023." https://www.statista.com/ - Statista https://www.statista.com/statistics/380270/nba-sponsorship-revenue/. Online; accessed 27 December 2023.

Statista (2023c). "National Basketball Association total league revenue from 2001/02 to 2022/23." https://www.statista.com/ - Statista https://www.statista.com/statistics/193467/total-league-revenue-of-the-nba-since-2005/. Online; accessed 27 December 2023.

Z. Terner and A. Franks (2021). "Modeling player and team performance in basketball." *Annual Review of Statistics and Its Application* **8**, 1–23.

R. Tunaru, E. Clark and H. Viney (2005). "An option pricing framework for valuation of football players." *Review of Financial Economics* **14**, 281–295. Real Options SI.

A. Wojnarowski (2022). "Mat Ishbia agrees to Suns purchase for record $4 billion." https://www.ESPN.com - Entertainment and Sports Programming Network. https://www.espn.com/nba/story/_/id/35292815/sources-mat-ishbia-finalizing-suns-purchase-4-billion. Online; accessed 30 June 2023.