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

Jackson P. Lautier*†

March 20, 2024

*Department of Mathematical Sciences, Bentley University
†Corresponding to jlautier@bentley.edu.

Abstract

An essential component of financial analysis is a comparison of realized returns. When both the inflows and outflows have dollar values, the calculations are rudimentary. Complexities arise if the returns are non-financial, however, such as on court basketball activities. To our knowledge, this problem remains open. We thus present the first known framework to estimate a return on investment for player salaries in the National Basketball Association (NBA). It is a five-part procedure that includes a novel basketball performance model, the WinLogit. The WinLogit is a per-game model that uses the relationship between team and individual statistics to assign fractional player credit to a team's win probability. The result is a wealth redistribution tool that allocates the revenue from a single game to each of its players. Using a player's salary as an initial investment, this creates a sequence of cash flows that may be evaluated using traditional financial analysis. The WinLogit is unbiased, calibrated to a replacement player, and we present its maximum likelihood estimate. The per-game approach allows for break-even analysis between high-performing players with frequent missed games and average-performing players with consistent availability. We illustrate all methods with empirical estimates from the 2022-2023 NBA regular season.

Keywords: internal rate of return, IRR, load management, player tracking data, PVWL, ROI

1 Introduction

Methods to assess the ongoing financial performance of invested monies are essential for financial analysts. Examples are ubiquitous: mutual fund fact sheets report historical returns, publicly-traded companies report quarterly earnings to shareholders, and lenders report on defaulted and delinquent loans. In the vast majority of these cases, both the cash inflows and outflows of invested capital may be recorded as market prices. This makes the financial return calculations rudimentary. Complexities arise when one side of the equation becomes non-financial, however. One such case is the player contract in the National Basketball Association (NBA). Namely, given a financial investment into an NBA player via a contractual salary, it is of interest to assess the realized return vis-à-vis on court activities (i.e., points, rebounds, etc.). It is not immediately clear how to value such on court performance in financial terms, and it is this curiosity that is the object of our study.

Such calculations would benefit numerous NBA stakeholders: e.g., informing player evaluations, informing roster building decisions, assessing team roster building competency, and comparing the relative financial efficiency of NBA teams and players. Furthermore, with the recent value of NBA franchises reaching $4 billion (Wojnarowski, 2022), the answers to these questions have become more important than ever. It is natural, then, to suppose there exists a great number of studies that consider both on court performance and salary simultaneously to arrive at methods to measure realized return on investment (ROI) or the internal rate of return (IRR) of a player's contract in view of said player's on court performance. A survey of related studies (e.g., Idson and Kahane, 2000; Berri et al., 2005; Tunaru et al., 2005; Berri and Krautmann, 2006; Berri et al., 2007a; Simmons and Berri, 2011; Halevy et al., 2012; Kuehn, 2017) indicates that this is not the case, however.

We thus propose the first known unified framework to consider both on court performance and salary concomitantly to derive a realized contractual ROI for players in the NBA. It is a five-part process that is summarized in Figure 1. The first step is to select a measurement period, such as a single NBA regular season. The next step is to determine a model to assign credit to players within a single game. We use a logistical regression model fitted to 36 individual player game data fields, including new player tracking data. This model utilizes the relationship between team statistics and individual player statistics to assign fractional player credit to team win probability (i.e., Theorem 2.1). The results of this model are then standardized and normalized to a replacement player analysis to derive a wealth redistribution model that rewards players with stronger relative performance and vice versa. We call this model a *WinLogit*, and it is itself a new entry into the growing field of basketball analytics. The WinLogit is unbiased (i.e., (13), Theorem 3.1) and calibrated to a replacement player (i.e., Theorem 2.2). We also find its maximum likelihood estimate (MLE) (i.e., Theorem 2.2). Further, its per-game approach allows for break-even calculations between high-performing players with frequent missed games and average-performing players with consistent availability (e.g., Figure 3).

The third step is to estimate a Single Game Value (SGV) in dollars. We create a model that tracks a SGV (i.e., revenue) in the form of attendance, television rights, and advertising revenue. All else equal, games that are nationally televised with greater attendance are more valuable. The fourth step is to combine the WinLogit and SGV to derive player cash flows that are based on relative on court performance. In other words, the WinLogit wealth redistribution model reassigns the SGV to each player in the game based on a system that rewards players who contribute more to winning with a higher share and vice versa. This completes the conversion from on court performance into a dollar value. Conditional on the WinLogit estimates, we demonstrate our wealth redistribution model is unbiased to total expected SGV (Theorem 3.1). The final step is to use a player's contractual salary as an invested cash flow and the now derived performance-based cash flows to solve for the ROI. The complete ROI process is summarized in Figure 1.

The paper proceeds as follows. Section 2 derives the WinLogit wealth redistribution model. Section 3 then builds upon the WinLogit to complete the ROI calculation. In both Sections 2 and 3, we provide empirical illustrations of all methods using data from the 2022-

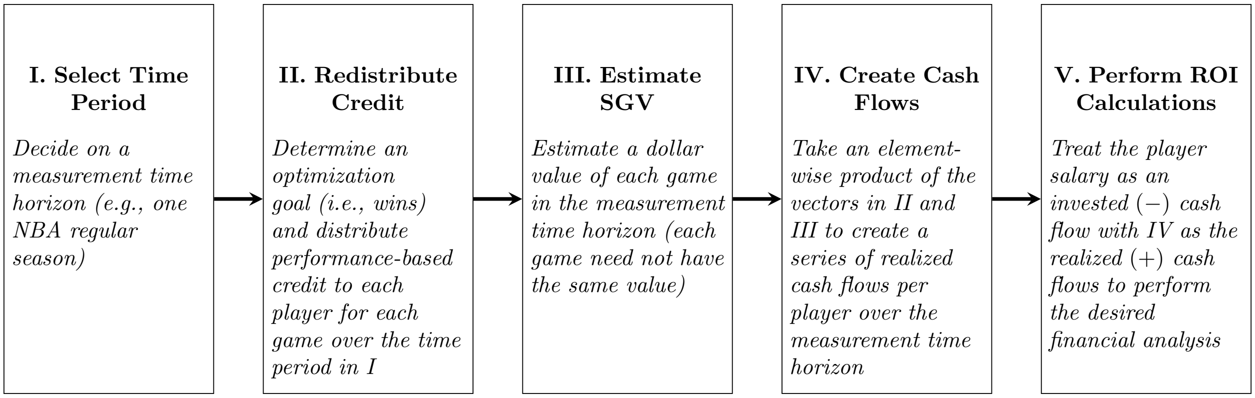

Figure 1: **NBA contractual ROI estimation framework summary**.

2023 NBA regular season. For comparison, we also provide empirical estimates of Game Score (Sports Reference LLC, 2023b) and a per-game version of Win Score (Berri et al., 2007b) throughout. The paper concludes in Section 4. The Appendix provides complete proofs, and the Supplemental Material includes a detailed literature review, model details, a robustness analysis, a simulation study, and an extension to Theorem 3.1. All data and replication code used herein may be found at https://github.com/jackson-lautier/nba_roi.

2 WinLogit

The purpose of the present section is to derive and illustrate a wealth redistribution tool to be used in part II of the ROI framework of Figure 1. It will proceed in two parts. First, Section 2.1 is the main methodological treatment, which culminates in the definition of the *WinLogit* in (10) and its statistical properties (Theorem 2.2). Next, we illustrate the WinLogit with data from the 2022-2023 NBA regular season in Section 2.2. Prior to this, we briefly review the existing literature to justify the need of the present section (a more detailed literature review may be found in the Supplemental Material).

Part II of the ROI framework of Figure 1 requires the basketball performance-based calculations to be contained within a single game unit. This is because the overall ROI framework of Figure 1 treats a player's contractual salary as invested capital that is intended to generate per game returns or positive payments. Particularly bad games become negative cash flows (losses), and missed games are treated as *defaults* or missed payments. Outside of the financial ROI framework of Figure 1, the purely basketball importance of the single game unit is well-known (e.g., Oliver, 2004, Chapter 16, pg. 192), and it is thus a natural delineation of NBA performance units. Furthermore, working on a per-game basis offers some advantages. For example, *per possession* standardization (e.g., Oliver, 2004, pg. 25) is not necessary because each team uses approximately the same number of possessions within one game (Berri et al., 2007b, pg. 101). From a statistical perspective, another advantage of a single game unit is that we may fit a logistic regression model, which can offer insights different than that of ordinary least squares (OLS). Finally, our per-game approach to performance measurement implies that running season per game totals (e.g., (15) of Section 2.2) allow analysts to determine the exact inflection point of a dominant player that misses many games versus a solid player that consistently plays (e.g., Figure 3.)

Does an existing model adequately meet our per-game requirements? Given what is available at present, we believe the answer is largely negative. Many previous studies have become dated when compared against recent player tracking data (e.g., Berri, 1999; Page et al., 2007; Fearnhead and Taylor, 2011; Martínez, 2012; Casals and Martínez, 2013). In a promising study, Lackritz and Horowitz (2021) create a model to assign fractional credit to scoring statistics for players in the NBA. Unfortunately, Lackritz and Horowitz (2021) consider only offensive statistics. Idson and Kahane (2000) and Tunaru et al. (2005) do not consider basketball. In a comprehensive review, Terner and Franks (2021) further our findings that a per-game approach is largely unstudied. Given these findings, Section 2.1 is itself novel within the context of basketball analysis literature.

One prevalent basketball performance statistic does limit all calculations to a single game: *Game Score* (Sports Reference LLC, 2023b). Per (Sports Reference LLC, 2023b), Game Score (GmSc) is defined as

$$\begin{aligned}\text{GmSc} = \text{PTS} + 0.4\text{FG} - 0.7\text{FGA} - 0.4(\text{FTA} - \text{FT}) \\ + 0.7\text{ORB} + 0.3\text{DRB} + \text{STL} + 0.7\text{AST} + 0.7\text{BLK} - 0.4\text{PF} - \text{TOV},\end{aligned} \tag{1}$$

where the abbreviations follow National Basketball Association (2023). Despite the per-game nature of (1), there are some limitations. First, GmSc does not utilize any of the recent NBA data advancements (National Basketball Association, 2023). Second, it relies on hard-coded coefficients, which are both difficult to interpret without greater context and potentially unstable over time. Finally, GmSc was derived outside of the peer-review process, which has garnered criticism (e.g., Berri and Bradbury, 2010).

Before proceeding to Section 2.1, we acknowledge there is a much discussed level of subjectivity to assigning credit to players in a basketball game (e.g., Oliver, 2004; Berri et al., 2007b). To this end, the WinLogit we propose does not need to be used within the ROI framework. Alternative per-game models, appropriately calibrated, may be swapped out for the WinLogit. For example, the Win Score (WSc) of Berri et al. (2007b), defined as

$$\begin{aligned}\text{WSc} = \text{PTS} + \text{ORB} + \text{DRB} + \text{STL} + 0.5\text{BLK} \\ + 0.5\text{AST} - \text{FGA} - 0.5\text{FTA} - \text{TOV} - 0.5\text{PF},\end{aligned} \tag{2}$$

may be instead recoded on a per-game basis. (As with (1), the abbreviations follow National Basketball Association (2023).) Because the intent of this manuscript is to provide an overall ROI framework design, of which the novel WinLogit we propose is only one component, we will reproduce all empirical results with (1) and a per-game version of (2) for comparison.

2.1 Methods

Deriving the WinLogit is a three-part process. The first step is to establish a set of principles to both calibrate the model and select input data. The next step is illustrating the merits of logistic regression within the context of basketball data (e.g., Theorem 2.1). Next, we translate the logistic regression model output into a wealth redistribution model. It is this last step that leads to the ultimate definition of the WinLogit; i.e., (10).

We employ three principles for data selection and model calibration: *Edwardsian* in outcome, value all activity, and no double counting. We now discuss each in turn.

Edwardsian in Outcome. We assume that NBA teams are *Edwardsian*; that is, NBA teams are attempting to maximize wins (Keeley, 2023) over the investment horizon. A wins-based objective function is quite standard in basketball analysis (e.g., Berri et al., 2007b, pg. 92). Other objective functions are possible, such as maximizing championships or maximizing operating income, see Section 4 for further discussion. Concisely, our logistic regression model is calibrated to win probability.

Value All Activity. From a classical statistics point-of-view, the model selection processes for exploratory observational studies often begins with data collection on a large scale (Kutner et al., 2005). As such, we desire to recognize any form of on court activity that has an effect on winning, both positive and negative. Pragmatically, this means that in addition to traditional box score categories, such as *two-point field goals made*, *turnovers*, and *blocks*, we also consider more recent player tracking and hustle statistics, such as *distance traveled*, *rebound chances*, *contested rebounds*, and *box outs*. This is an advantage of using new player tracking data in comparison to (1) and (2), though the trade-off is added complexity. In addition to data collection, we also consider this principle is selecting a logistic regression model. Specifically, we desire to recognize players with strong games despite losing at the team level. Hence, our model allows a player to make a positive individual contribution to win probability despite poor team play overall and vice versa. As a minor comment, we are at times constrained by data availability (e.g., it is preferable to track "screens set" instead of *screen assists*, but detailed data for screens set by game is not yet readily available).

No Double Counting. We desire to avoid the classic economics problem of *double counting*, which is undesirable in the measurement of macroeconomic calculations like *gross domestic product* (e.g., Mankiw, 2003, Chapter 10). In essence, our objective is to avoid giving a player double credit. For example, we create statistics such as three-point field goals missed rather than use both three-point field goals made and three-point field goal attempts. Similarly, we track two-point field goals made, three-point field goals made, and free throws made but do not also track total points scored. Other nonobvious adjustments include subtracting rebounds from *rebound chances*, subtracting blocks from *contested two-point shots*, subtracting *charges drawn* from *personal fouls drawn*, and subtracting assists, *secondary assists*, and *free throw assists* from *passes made*. In reviewing (1) and (2), we see that each equation tracks both field goals (FG) or points (PTS) and field goals attempted (FGA), which would violate this principle. Hence, the WinLogit approach may offer a novel economic perspective that differs from these traditional basketball measures.

From these three principles, our initial data collection consists of 36 player-level statistical categories: made two-point shots (FG2O), missed two-point shots (FG2X), made three-point shots (FG3O), missed three-point shots (FG3X), made free throws (FTMO), missed free throws (FTMX), personal fouls (PF), steals (STL), adjusted offensive rebounds (i.e., offensive rebounds less contested offensive rebounds) (AORB), adjusted defensive rebounds (ADRB), assists (AST), blocks (BLKS), turnovers (TO), blocks against (BLKA), adjusted personal fouls drawn (i.e., personal fouls drawn less charges drawn) (PFD), screen assists (SAST), deflections (DEFL), charges drawn (CHGD), adjusted contested two-point shots (i.e., contested two-point shots less blocks) (AC2P), contested three-point shots (C3P), offensive box outs (OBOX), defensive box outs (DBOX), offensive loose balls recovered (OLBR), defensive loose balls recovered (DLBR), defended field goals against made (DFGO), defended field goals against missed (DFGX), drives (DRV), distance traveled in miles offense (ODIS), distance traveled in miles defense (DDIS), adjusted passes made (i.e., passes made less assists, secondary assists, and free throw assists) (APM), secondary assists (AST2), free throw assists (FAST), offensive contested rebounds (OCRB), defensive contested rebounds (DCRB), adjusted offensive rebound chances (i.e., offensive rebound chances less offensive rebounds) (AORC), and adjusted defensive rebound chances (ADRC). All adjustments are made to align with the *No Double Counting* principle. For complete definitions of these fields, see National Basketball Association (2023).

To perform the statistical analysis, we will employ a logistic regression model as follows (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 is a Bernoulli random variable with parameter, p_i, for $i = 1, \ldots, n$. That the model is attuned to the *Edwardsian in Outcome* principle is immediate. The binary logit regression model has the form, for $i = 1, \ldots, n$,

$$f(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) = \frac{\exp(y_i \boldsymbol{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}, \tag{3}$$

or,

$$\operatorname{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{x}_i^\top \boldsymbol{\beta}. \tag{4}$$

The form (4) implies

$$p_i = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})} = \frac{1}{1 + \exp(-\boldsymbol{x}_i^\top \boldsymbol{\beta})}. \tag{5}$$

Hence, the regression coefficients are called log-odds ratios. That is, 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, $j, j^* = 1, \ldots, k$. Thus, at the team level, any field in $\mathbf{X}$ that returns a positive (and significant) coefficient estimate can be interpreted as having a positive contribution to winning and vice versa for negative coefficients (i.e., *Edwardsian in Outcome*).

Logistic regression in the context of basketball game data 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 $p(\bar{X}_1, \ldots, \bar{X}_k) = \exp(\beta_0)(1+\exp(\beta_0))^{-1}$ probability of winning. Hence, $\beta_0 = 0$ implies $p(\bar{X}_1, \ldots, \bar{X}_k) = 0.5$, a quite reasonable assumption. Second, if we both assume $\beta_0 = 0$ and that each NBA team has the required roster of 15 players per game (National Basketball Association, 2018), then we may distribute the logit of the team's win probability linearly to the logit of each player's individual win probability. This is a direct result of team level statistics equaling the sum of individual player level statistics (one minor exception is that a team turnover is not credited to an individual player). We formalize this desirable property in Theorem 2.1.

Theorem 2.1. *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

$$X_{ij\bullet} - \bar{X}_{ij\bullet} = \sum_{m=1}^{15} \Big(X_{ijm} - \bar{X}_{ijm} \Big), \tag{6}$$

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 (4), then

$$\text{logit}(p_i) = \sum_{m=1}^{15} \text{logit}(p_{im}), \tag{7}$$

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.

Proof. See Appendix A. □

The first part of Theorem 2.1 may be reminiscent of finding the treatment effects of balanced experiment designs (e.g., Montgomery, 2020, §3.3.3).

Finally, it is left to translate the player level game logit into a fractional share of the entire game. Because $\text{logit}(p_{im})$ may be negative, this task requires careful consideration beyond a traditional percentage calculation. Note that both the GmSc and WSc calculations may also be negative for a single game, and so both (1) and (2) require the same careful consideration. Recall that a property of the logistic model in (3) with centered covariates and $\beta_0 = 0$ is that an average game in all statistical categories for a player m yields $p_{im} = 0.5$ or $\text{logit}(p_{im}) = 0$, for any game outcome i, $i = 1, \ldots, n$. Hence, $\text{logit}(p_{im}) > 0$ suggests an "above average" game, while $\text{logit}(p_{im}) < 0$ suggests a "below average" game. Further, Theorem 2.1 suggests that the team level logit follows the same interpretation. In other words, we can imagine that both teams within a single game are competing to obtain the largest team logit, with individual players making both positive and negative contributions.

Because we desire to compute ROI calculations from on court performance only, we will restrict all subsequent measures to the set of players with playing time over the investment horizon. From this perspective, we assume that a game is worth on average one unit (i.e., one win) and that all players with playing time (i.e., $\texttt{minutes} > 0$), denoted $\mathcal{M}_g$ for game g, $g = 1, \ldots, n/2$, begin with a $1/\bar{m}$ share, where $\bar{m} = m^*/(n/2)$ and m^* is the total number of players with playing time in the $n/2$ total games (i.e., $m^* = \sum_g \sum_{m \in \mathcal{M}_g} 1$). For ease of interpretation, we desire that an average game results in the same $1/\bar{m}$ share. Further, we prefer the measure to be standardized for ease of comparison. Hence, define the basic sample statistics

$$\overline{\text{WL}}_{m^*}(\boldsymbol{\beta}) = \frac{1}{m^*} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{logit}(p_{gm}), \tag{8}$$

and

$$s(\text{WL})_{m^*}(\boldsymbol{\beta}) = \sqrt{\frac{1}{m^* - 1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(\text{logit}(p_{gm}) - \overline{\text{WL}}_{m^*} \right)^2}. \tag{9}$$

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

WinLogit_{gm}, as

$$\text{WinLogit}_{gm}(\boldsymbol{\beta}) = \frac{1}{s(\text{WL})_{m^*}}\left(\text{logit}(p_{gm}) - \overline{\text{WL}}_{m^*}\right)\frac{1}{\bar{m}} + \frac{1}{\bar{m}}. \tag{10}$$

The form of (10) suggests that below average games decreases a player's share, and above average games increases a player's share. Hence, (10) accounts the fundamental *replacement player* adjustment widely preferred across sports (e.g., Shea and Baker, 2012). Pleasingly, (10) allows players on a losing team that have a strong game to still receive a positive share of the game's value and vice versa. This aligns with the *Value all Activity* principle. The appearance of $\boldsymbol{\beta}$ in the build up to (10) is to remind us that the WinLogit is a function of the parameters defined in (3) through (5). The WinLogit has some attractive statistical properties, which we now summarize.

Theorem 2.2. *Let the* WinLogit_{gm} *take the form of* (10) *for player* $m \in \mathcal{M}_g$, $g = 1, \dots, n/2$. *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 $\hat{\boldsymbol{\beta}}_{MLE}$ *be the MLE of the logistic regression assumed in Theorem 2.1. Then the MLE of* $\text{WinLogit}_{gm}(\boldsymbol{\beta})$ *is* $\text{WinLogit}_{gm}(\hat{\boldsymbol{\beta}}_{MLE})$.

Proof. See Appendix A. □

In an economic interpretation, the WinLogit may be thought of as a 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. For the sake of equivalent comparison, we may also use (1) to define for player

$m \in \mathcal{M}_g$ in game g, $g = 1, \ldots, n/2$

$$\text{GmSc}^*_{gm} = \frac{1}{s(\text{GS})_{m^*}} \left(\text{GmSc}_{gm} - \overline{\text{GS}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}, \tag{11}$$

where $\overline{\text{GS}}_{m^*} = \frac{1}{m^*} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{GmSc}_{gm}$ and $s(\text{GS})^2_{m^*} = \frac{1}{m^*-1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} (\text{GmSc}_{gm} - \overline{\text{GS}}_{m^*})^2$. Similarly, via (2) we define for player $m \in \mathcal{M}_g$ in game g, $g = 1, \ldots, n/2$

$$\text{WnSc}^*_{gm} = \frac{1}{s(\text{WS})_{m^*}} \left(\text{WnSc}_{gm} - \overline{\text{WS}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}, \tag{12}$$

where $\overline{\text{WS}}_{m^*} = \frac{1}{m^*} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{WnSc}_{gm}$ and $s(\text{WS})^2_{m^*} = \frac{1}{m^*-1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} (\text{WnSc}_{gm} - \overline{\text{WS}}_{m^*})^2$. Both (11) and (12) preserve the standardization of Theorem 2.2. Hence, we can directly compare wealth allocation differences between (10), (11), and (12) in the sequel (e.g., Figure 2). Finally, observe that by definition

$$\sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{WinLogit}_{gm} = \frac{n}{2}, \tag{13}$$

which ensures an unbiased estimate at the aggregate level (i.e., the total reallocation of wins must sum to the original total of wins, $n/2$). The property in (13) holds for both (11) and (12), as well.

2.2 Empirical Results

We now employ the methods of Section 2.1 to NBA player statistics from the 2022-2023 NBA regular season (National Basketball Association, 2023). To compile the necessary statistics, we utilize the `python` package `nba_api` (Patel, 2018). Because we require game-by-game statistics, we design a custom game-by-game query wrapper for Patel (2018). The result is a novel data set of 1,226 2022-2023 NBA regular season games (i.e., n = 2,452) spanning all 36 statistical categories described in Section 2.1. Four games did not report player tracking data and were excluded: GSW @ SAS on January 13, 2023, CHI @ DET on January 19, 2023, POR @ SAS on April 6, 2023, and MIN @ SAS on April 8, 2023. To obtain the data and replication code, please navigate to the public `github` repository at https://github.com/jackson-lautier/nba_roi.

We first fit a logistic regression model at the team level for all 36 statistical categories identified in Section 2.1. We then remove covariates that are not statistically significant at $\alpha = 0.10$: BLKA, AST, DEFL, OBOX, OLBR, DLBR, DRV, ODIS, AST2, FAST, AORC, and ADRC and perform a second logistical regression. In the second model, we estimate $\hat{\beta}_0 = -0.004930$ with a p-value of 0.948. Hence, we may comfortably refit the logistical regression without an intercept, as it only results in a negligible amount of bias. Because we may use Theorem 2.1 with $\beta_0 = 0$, we feel allowing such small estimation bias is a worthwhile trade-off (specifically, the total estimated win probability under the $\beta_0 = 0$ model is 1,226.88, and the unbiased total would be 1,226, half of the sample size; further, the form of (10) will correct this bias per (13)). A summary of the fitted model may be found in Table 1. The Supplemental Material contains the initial model fitting parameters with all 36 data fields.

The model of Table 1 suggests that missing shots (i.e., FG2X, FG3X, FTMX), committing fouls (PF) and turnovers (TOV), contesting three point shots (C3P), allowing baskets on defended shots (DFGO), and defensive distance traveled (DDIS) negatively impact win probability. Of these, the only surprise is C3P, though it may be highly related to opponents making three point shots. On the winning side, it is beneficial to make baskets (i.e., FG2O, FG3O, FTMO), collect rebounds (AORB, ADRB), steals (STL), blocks (BLK), draw non-charge fouls (PFD), draw charges (CHGD), set screen assists (SAST), contest two-point shots (AC2P), box out on the defensive end (DBOX), have contested shots miss (DFGX), make passes not counted in assists (APM), and collect contested rebounds (OCRB, DCRB). The most important statistical categories may be assessed by a standard variable importance analysis (Kuhn, 2008). It finds that making (FG3O) and missing (FG3X) three-point field goals are the most important determinants of winning. This aligns closely with long-term trend analysis of the NBA (e.g., Goldsberry, 2019). There are apparent disagreements be-

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 1: **WinLogit Logistic Regression Model Parameters**. Based on team outcomes for the 2022-2023 NBA regular season. Because player tracking data was not available for four games, 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).

tween the model of Table 1, (1), and (2). One example is that the model of Table 1 does not report assists (AST) as significant but instead finds adjusted passes made (APM) as significant. Contrast this to (1) and (2), both of which use assists. This is perhaps a result of using player tracking data, which allows for more detail than either (1) or (2).

Returning to the wealth redistribution interpretation, we may compare the resulting distributions of (10), (11), and (12) in Figure 2. We see that despite having the same mean and standard deviation of $1/\bar{m} = 4.75\%$, the distributions differ. Specifically, the WinLogit is more symmetric, whereas both the Game Score and Win Score are skewed right. Furthermore, we may assess the cumulative total performance of a player for the entire regular season. To do so, let $\mathcal{G}_m$ represent the set of games of which player m's team appeared (typically $\#\{\mathcal{G}_m\} = 82$ for a standard NBA regular season). The set $\mathcal{G}_m$ may be larger than the set of games for which player m recorded playing time due to injuries or coaching decisions. Hence, define for any $g \in \mathcal{G}_m$,

$$\text{WinLogit}^*_{gm} = \begin{cases} \text{WinLogit}_{gm}, & m \in \mathcal{M}_g \\ 0, & m \notin \mathcal{M}_g. \end{cases} \tag{14}$$

Because $\sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \text{WinLogit}^*_{gm} = n/2$ still holds trivially, the desirable unbiased property of (13) remains. In financial parlance, the form of (14) implies a missed game is a *default*. The season total of (14) for player m is then

$$\text{PVWL}_m = \sum_{g \in \mathcal{G}_m} \text{WinLogit}^*_{gm}. \tag{15}$$

We may consider (15) as a present value of a series of cash flows taking the value of (14) discounted at a zero interest rate. In other words, (15) assumes all single game values are unity. This allows for a pure performance measure that does not include salary. Notably, the game-by-game approach including zeros used in (14) allows for an instant comparison of a high-performing player with frequent missed games against an average-performing player with consistent availability (i.e., Figure 3). This has been a source of perturbation in evaluating players among NBA pundits (e.g., Lowe, 2020), of which (15) may offer new insights.

By (15), the top five PVWL performers for the 2022-2023 NBA regular season relative to position (player position data per RealGM, L.L.C. (2023); Sports Reference LLC (2023a)) are (1) Luka Dončić (PG, 66 GP, 6.159 PVWL), (2) Jayson Tatum (SF, 74 GP, 6.707 PVWL), (3) Giannis Antetokounmpo (PF, 63 GP, 8.096 PVWL), (4) Shai Gilgeous-Alexander (PG, 68 GP, 5.036 PVWL), and (5) Nikola Jokić (C, 69 GP, 10.088 PVWL).[*] Without correcting for position, Nikola Jokic is the top overall PVWL performer. We may do the same but replace the WinLogit with WnSc^*_{gm} and GmSc^*_{gm}. (In what follows, we employ the same

[*]The standard position abbreviations are point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C). The abbreviation GP denotes *games played*.

adjustment of (14) to WnSc^*_{gm} and GmSc^*_{gm} but abuse notation for ease of exposition.) Specifically, for $\text{PVWS}_m = \sum_{g \in \mathcal{G}_m} \text{WnSc}^*_{gm}$ for player m, the top five performers for the 2022-2023 NBA regular season relative to position are (1) Jayson Tatum (SF, 74 GP, 8.238 PVWS), (2) Luka Dončić (PG, 66 GP, 8.025 PVWS), (3) Nikola Jokić (C, 69 GP, 11.505 PVWL), (4) Domantas Sabonis (C, 79, 11.016 PVWL), and (5) Giannis Antetokounmpo (PF, 63 GP, 7.905 PVWL). Without correcting for position, Nikola Jokic is also the top overall PVWS performer. From the the perspective of $\text{PVGS}_m = \sum_{g \in \mathcal{G}_m} \text{GmSc}^*_{gm}$ for player m, the top five performers for the 2022-2023 NBA regular season relative to position are (1) Jayson Tatum (SF, 74 GP, 9.785 PVGS), (2) Nikola Jokić (C, 69 GP, 10.426 PVGS), (3) Joel Embiid (C, 66 GP, 10.331 PVGS), (4) Donovan Mitchell (SG, 68 GP, 7.990 PVGS), and (5) Giannis Antetokounmpo (PF, 63 GP, 8.872 PVGS). Without correcting for position, Nikola Jokic is also the top overall PVGS performer. For reference, because $1/\bar{m} = 4.75\%$, an average player playing 82 games would obtain a PVWL, PVWS, or PVGS of 3.896.

There is general agreement between all three methods (i.e., Jayson Tatum, Nikola Jokić, and Giannis Antetokounmpo appear on all three top 5 lists, and all have Nikola Jokić as the top overall performer, irrespective of position). Despite the relative agreement of the top performers, there are notable model differences. For a summary of the top disagreements between sum totals of (10), (11), and (12) along the lines of (15), see Table 2. The robustness analysis of the Supplemental Material finds the WinLogit produces the smallest errors in projecting team wins (absolute total and team rank) and is itself the most significant in predicting win probability among the three methods considered. For complete results, navigate to the public `github` repository at https://github.com/jackson-lautier/nba_roi.

3 Return on Investment

The purpose of the present section is to complete parts III, IV, and V of the ROI framework of Figure 1. This part of the paper builds upon the earlier statistical analysis towards

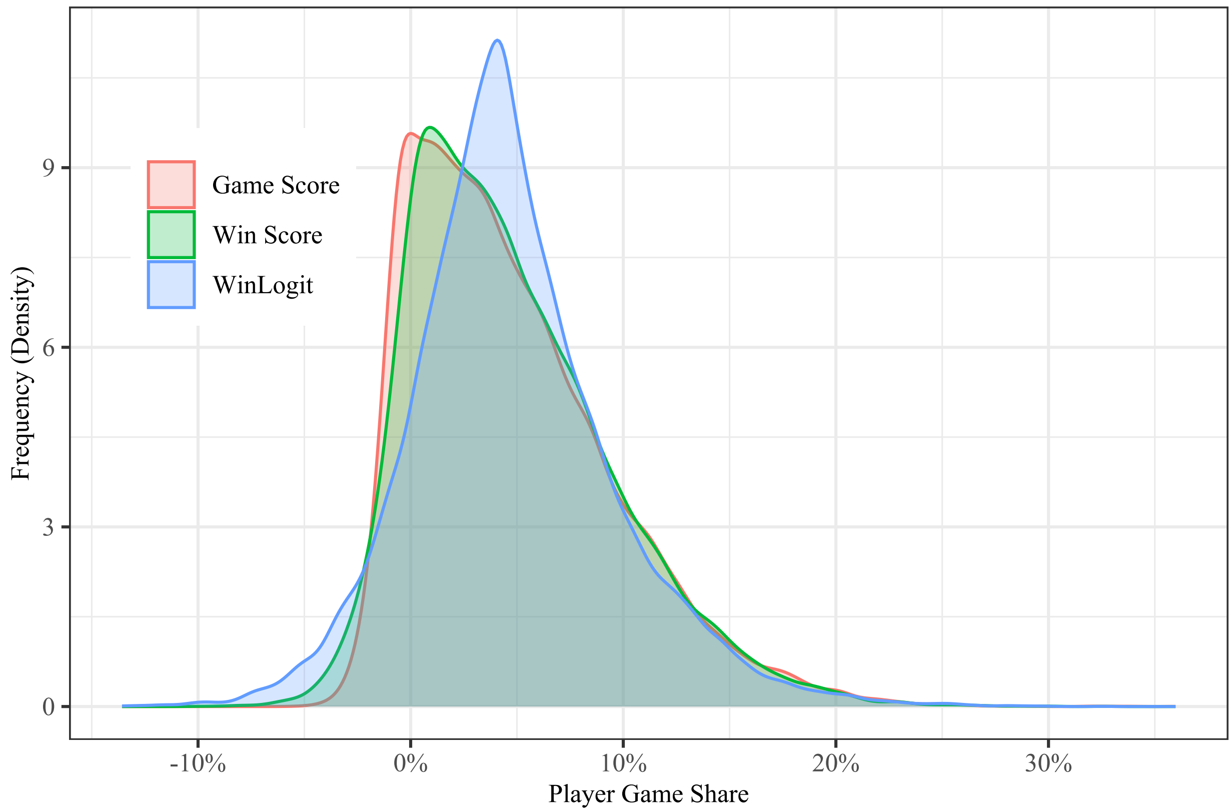

Figure 2: **Wealth redistribution comparison**. Frequency distributions of (10), (11), and (12) for all NBA players from the 2022-2023 NBA regular season. The sample of n = 2,452 games includes m^* = 25,804 individual players with playing time.

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 2: **Player performance disagreements**. The top ten largest disagreements between sum totals of (10), (11), and (12) for the 2022-2023 NBA regular season in terms of percentile rank (%).

the financial methods and results. The section proceeds in two parts. First, Section 3.1 introduces a model for the Single Game Value (part III) and an unbiased technique to create the cash flows (part IV). It also reviews how to calculate an ROI once the cash flows have been modeled (part V). Next, Section 3.2 illustrates our ROI framework with data from the

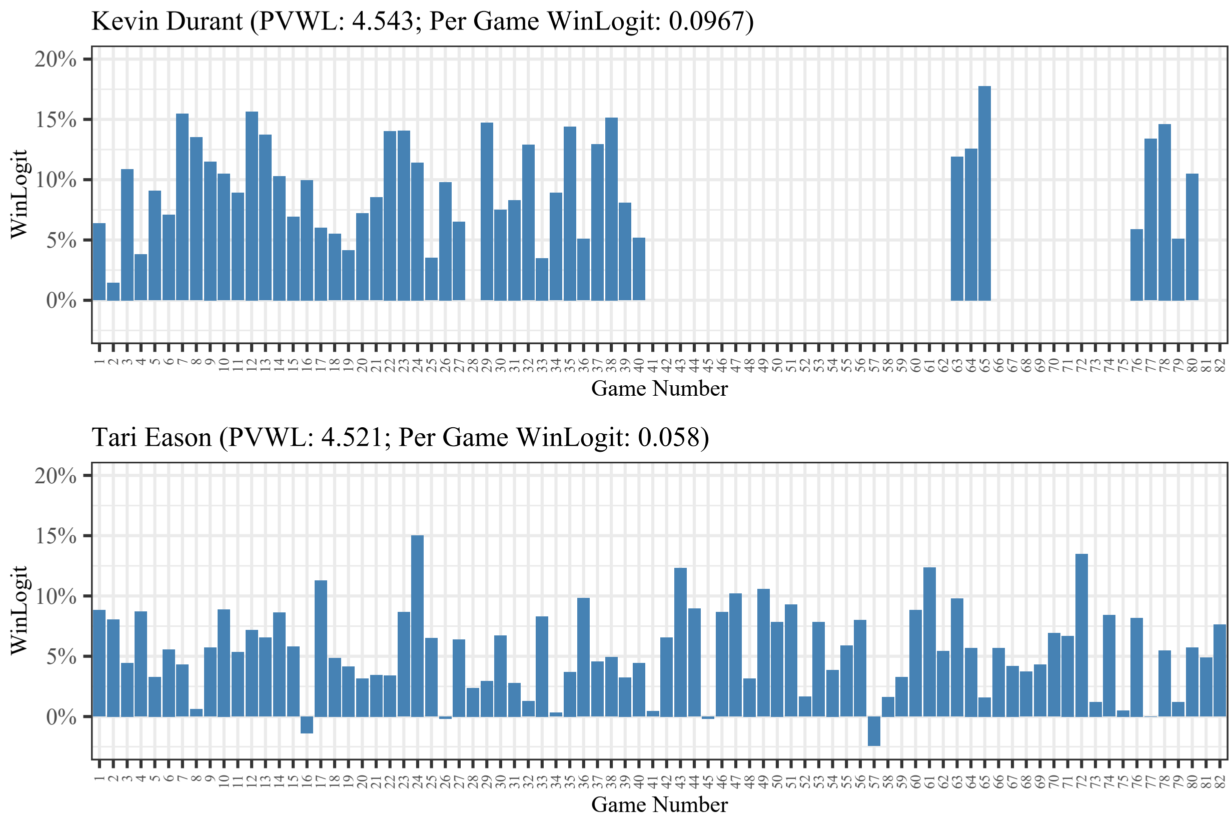

Figure 3: **Quantifying missed games**. The per-game approach of (15) allows for break-even calculations between high-performing players with frequent missed games (Kevin Durant, 47 games played, top) against average-performing players with consistent availability (Tari Eason, 82 games played, bottom). Data spans the 2022-2023 NBA regular season.

2022-2023 NBA regular season. Prior to this, we briefly review the related literature (the Supplemental Material provides a more detailed literature review).

While no studies consider both player salary and on court performance simultaneously, there is related work outside of basketball (e.g., Idson and Kahane, 2000; Tunaru et al., 2005). The field of sports economics within basketball considers competitive imbalances (Berri et al., 2005), shirking (Berri and Krautmann, 2006), and salaries (Berri et al., 2007a; Simmons and Berri, 2011; Halevy et al., 2012; Kuehn, 2017). Our forthcoming analysis differs from all of these studies generally in that we do not attempt to explain salary decisions. Instead, we propose the first known framework to measure the realized return of a player's contract in light of on court performance.

3.1 Methods

It remains to estimate the SGV (step III), derive the performance-based cash flows (step IV), and perform the ROI calculations (step V) to complete the ROI framework of Figure 1. Specifically, we first propose a method to model the SGV. We then briefly review how to perform a standard financial ROI calculation from a sequence of cash flows. Next, we use the SGV model and the results of Section 2.1 to derive an unbiased estimate of a player's performance-based cash flows. Finally, we combined everything into an optimization function, the solution of which is a player's ROI estimate.

Modeling a SGV is equivalent to answering the question: how does a regular season NBA game generate revenue? Variations of this question have attracted previous attention (e.g., Berri et al., 2007b, Chapter 5). In working from the basic ideas of Berri et al. (2007b), we assume NBA revenue is generated from ticket sales and television rights. We add a third component, which is revenue from advertising. Specifically, for $g = 1, \ldots, n/2$, define the parametric random variable

$$\mathrm{SGV}_g(\boldsymbol{\alpha}) = \alpha_1 \mathrm{GATE}_g + \alpha_2 \mathbf{1}_{\mathrm{ESPN}} + \alpha_3 \mathbf{1}_{\mathrm{TNT}} + \alpha_4(\mathbf{1}_{\mathrm{ESPN}} + \mathbf{1}_{\mathrm{TNT}} + \mathbf{1}_{\mathrm{NBATV}}), \qquad (16)$$

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. In proposing (16), we do not assume a game televised on NBATV generates television rights revenue for the NBA, but we do assume it generates advertising revenue.

In words, we propose to model SGV_g as the sum total of ticket sales, television revenue, and advertising revenue from game g, $g = 1, \ldots, n/2$. The natural assumption is that games with higher attendance will be worth more, all else equal, and games that are nationally televised will be worth more, all else equal. This allows us to approximate the relative importance of a game, and it results in the intuitive outcome that players with more nationally televised games will generate a better ROI. This latter point connects with previous studies, as part of the value of signing star players is greater attention from fans and advertisers (e.g., Berri et al., 2007b, Chapter 5). This approach does not consider a team's relative position in the standings, which is an alternative perspective on a game's importance. This is a potentially meaningful modeling decision, which we discuss more in Section 4.

With an approach to model the SGVs in hand, we may move to deriving the performance-based cash flows (i.e., step IV in Figure 1). Before doing so, it is instructive to review how to calculate the realized ROI for a sequence of financial cash flows generally. We will utilize the *internal rate of return* methodology of Berk and Demarzo (2007, §4.8). Let CF_0 be the initial (i.e., negative) investment, and $\mathrm{CF}_1, \ldots, \mathrm{CF}_N$ be the positive future cash flows. For simplicity, we assume all cash flows occur on equally spaced intervals. Because we are performing a realized, ex post, return calculation, all CF_t, $t = 1, \ldots N$, are assumed known. The return on investment is the rate, r, such that

$$\mathrm{CF}_0 = \sum_{t=1}^{N} \frac{\mathrm{CF}_t}{(1+r)^t}. \tag{17}$$

Aside from very simplified versions of (17), the computation of r will require the use of optimization software (e.g., Varma, 2021).

To utilize (17) within the context of the NBA ROI modeling framework we propose, therefore, it is left to derive the cash flows. To do so, we first assume the time zero cash flow (i.e., CF_0) is a player's full salary over the investment time horizon and is paid in a single lump sum. For example, assuming an NBA regular season, CF_0 would represent a full season salary. From the perspective of the NBA team, it is a negative cash flow and represents the initial investment. To find the return cash flows, CF_t, $t = 1, \ldots N$, we may use any of (10), (11), and (12) in conjunction with (16). For ease of exposition, we shall assume (10).

Formally, for any distinct player $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$, let $\mathbf{SGV}_{g \in \mathcal{G}_m} = (\text{SGV}_1, \ldots, \text{SGV}_N)^\top$ be a vector of SGVs, via (16), 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 $\mathbf{WL}_{g \in \mathcal{G}_m} = (\text{WinLogit}^*_{1m}, \ldots, \text{WinLogit}^*_{Nm})^\top$ be a vector of WinLogits, via (14), 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

$$\mathbf{CF}_m = (\mathbf{SGV}_{g \in \mathcal{G}_m})^\top \text{diag}(\mathbf{WL}_{g \in \mathcal{G}_m}) = (\text{SGV}_1 \text{WinLogit}^*_{1m}, \ldots, \text{SGV}_N \text{WinLogit}^*_{Nm})^\top, \quad (18)$$

where $\text{diag}(\mathbf{WL}_{g \in \mathcal{G}_m})$ represents a diagonal $N \times N$ matrix with diagonal $\mathbf{WL}_{g \in \mathcal{G}_m}$. By the definition of (10), it is possible a particularly bad game may result in $\text{SGV}_t \text{WinLogit}^*_{tm} < 0$ for some t, $t = 1, \ldots, N$ and player $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$.

Before proceeding to complete the ROI methodology, we illustrate that the form (18) has a desirable conditional unbiasedness property. Specifically, recall that (10) may be thought of as a wealth redistribution model that reallocates the SGV based on a player's on court performance. Hence, it is of interest to ensure the reallocated cash flows in (18), given a fitted model in (10), are unbiased to the expected sum total of all SGVs, i.e., $\mathbf{E}(\sum_{g=1}^{n/2} \text{SGV}_g)$. In other words, we do not wish to inadvertently "create" or "eliminate" wealth due to a faulty model. This property holds if $\mathbf{E}(\text{SGV}_g) = \mu \in \mathbb{R}$ for all $g = 1, \ldots, n/2$, which we now show.

Theorem 3.1. *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 (10)*, when viewed over all players and games in the investment time horizon, is unbiased to the expected total generated revenue.*

Proof. See Appendix A. □

Theorem 3.1 will hold for (16), though it is a more general result. Finally, to retrieve the form of (17), 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 $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$ is distinct. Then the contractual ROI for distinct player $m \in \{\mathcal{M}_g\}_{1 \leq g \leq n/2}$ over the investment time horizon, is the rate, r_m, such that

$$\mathrm{CF}_0^m = (\mathbf{SGV}_{g \in \mathcal{G}_m})^\top \mathrm{diag}(\mathbf{WL}_{g \in \mathcal{G}_m}) \boldsymbol{\nu}_m = \sum_{t=1}^{N} \frac{\mathrm{SGV}_t \mathrm{WinLogit}^*_{tm}}{(1 + r_m)^t}, \tag{19}$$

where CF_0^m is distinct player m's full salary over the investment time horizon. This is the last and final step to complete the ROI framework of Figure 1. We remark that (19) relies on a set of reasonable assumptions, which are discussed more fully in Section 4.

3.2 Empirical Results

We first estimate the parameters of (16) before proceeding to the ROI calculations. As in Section 2.2, all results correspond to the 2022-2023 NBA regular season. Our estimates rely on various data sources and proceed as follows.

Attendance figures are readily available per game (e.g., National Basketball Association, 2023), which allows for a reliable estimate of GATE_g, $g = 1, \ldots, n/2$. To estimate α_1, we may work backwards from total NBA revenue. Specifically, total gates for the 2022-2023 NBA regular season are known to be 21.57% of total NBA revenue (Statista, 2023a). Further, total NBA revenue for the 2022-2023 NBA regular season is known to be \$10.58B (Statista, 2023c). Hence, we may estimate total gate revenue at $\$10.58 \times 21.57\% = \$2.28$B. With total attendance for the 2022-2023 NBA regular season at 22,234,502 (National Basketball Association, 2023), we arrive at an estimate of the average per-ticket price, $\hat{\alpha}_1 = \$102.64$.

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 3: **Component Estimates of $\mathbf{SGV}_g$.** Coefficient estimates of (16) based on available data for the 2022-2023 NBA regular season (National Basketball Association, 2023; Statista, 2023a,c; Lewis, 2023; Statista, 2023b).

To estimate α_2, α_3, and α_4, we may again work backwards from total NBA revenue. Specifically, it is known that total NBA television revenue for the 2022-2023 NBA regular season is \$1.4B for games televised on ESPN (Lewis, 2023) and \$1.2B for games televised on TNT (Lewis, 2023). With 101 games televised on ESPN (National Basketball Association, 2023) and 65 games televised on TNT, we estimate $\hat{\alpha}_2 = \$13{,}861{,}386$ and $\hat{\alpha}_3 = \$18{,}461{,}538$. Finally, total NBA advertising revenue for the 2022-2023 NBA regular season is known to be \$1.66B (Statista, 2023b). As an approximation, we assume total ad revenue to be spread equally among the 273 nationally televised 2022-2023 NBA regular season games (ESPN: 101; TNT: 65; NBATV: 107) (National Basketball Association, 2023). Hence, we estimate $\hat{\alpha}_4 = \$6{,}080{,}586$. A summary of coefficient estimates for (16) may be found in Table 3. For reference, the top five teams in terms of total SGV for the 2022-2023 NBA regular season are LAL (\$908.3M), GSW (\$885.4M), BOS (\$831.1M), PHX (\$766.3M), and PHI (\$708.5M). Each of these teams play in some of the largest television media markets (Sports Media Watch, 2024), which helps to validate these estimates. Players on these teams will generate higher ROIs because the games are more valuable, all else equal.

To estimate contractual ROI, we obtain salary data for all players from the 2022-2023 NBA regular season from HoopsHype (2023) (with one supplement for the player Chance Comanche (Spotrac, 2023)). Therefore, with the estimates in Table 3 and the earlier work of Section 2.2, we are able to perform the ROI calculations using (19). The results, assuming a minimum games played of 42, are as follows.

By (14), the top five ROI performers for the 2022-2023 NBA regular season relative to position (player position data per RealGM, L.L.C. (2023); Sports Reference LLC (2023a)) are (1) Santi Aldama (PF, \$2.09M SAL, 53.21% ROI), (2) John Konchar (SF, \$2.30M SAL, 41.78% ROI), (3) Jock Landale (C, \$1.56M SAL, 32.96% ROI), (4) Austin Reaves (SG, \$1.56M SAL, 33.24% ROI), and (5) Jose Alvarado (PG, \$1.56M SAL, 16.43% ROI). Without correcting for position, Santi Aldama is the top overall ROI contract. We may do the same but replace (14) with WnSc^*_{gm} and GmSc^*_{gm}. (In what follows, we employ the same adjustment of (14) to WnSc^*_{gm} and GmSc^*_{gm} but abuse notation for ease of exposition.) Specifically, for WnSc*, the top five performers for the 2022-2023 NBA regular season relative to position are (1) Santi Aldama (PF, \$2.09M SAL, 51.98% ROI), (2) John Konchar (SF, \$2.30M SAL, 36.08% ROI), (3) Jose Alvarado (PG, \$1.56M SAL, 16.96% ROI), (4) Jock Landale (C, \$1.56M SAL, 22.46% ROI), and (5) Austin Reaves (SG, \$1.56M SAL, 19.90% ROI). Without correcting for position, Santi Aldama is also the top overall ROI contract under the WnSc* wealth redistribution method. From the the perspective of GmSc*, the top five performers for the 2022-2023 NBA regular season relative to position are (1) Santi Aldama (PF, \$2.09M SAL, 29.12% ROI), (2) Jock Landale (C, \$1.56M SAL, 25.70% ROI) (3) Tyrese Maxey (SG, \$2.73M SAL, 32.19% ROI), (4) Jose Alvarado (PG, \$1.56M SAL, 19.70% ROI), and (5) Naji Marshall (SF, \$1.78M SAL, 20.34% ROI). Without correcting for position, Tyrese Maxey is the top overall ROI contract under the GmSc* wealth redistribution method. In terms of relative performance by position, there is general agreement between all three methods. Santi Aldama, Jock Landale, and Jose Alvarado appear on all three top 5 relative to position lists. Further, both (14) and WnSc* find Santi Aldama as the player with the top overall ROI contract, while GmSc* finds Tyrese Maxey to be the top overall ROI contract.

We may also use traditional financial calculations to compare the risk-reward by position. For example, the *coefficient of variation* (CV) (Klugman et al., 2012, Definition 3.2, pg. 20) takes a ratio of the standard deviation of an asset class to its mean. Hence, if we consider each position as an asset class, we may perform the same calculation. We do so in Table 4. Both WinLogit* and WnSc* find the center position offers the least variability in 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 4: **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.

relative to the mean return. Conversely, GmSc* suggests the small forward (SF) position offers the least variability in return relative to the mean. Further, WinLogit*-based ROIs shown large risk-return differences by position, whereas GmSc*-based ROIs show CVs that are much closer together. For reference, we may calculate a replacement player ROI. Recall we have normalized (10), (11), and (12) to $1/\bar{m} = 4.75\%$. Further, we obtain an average SGV of $5,318,785, which yields a replacement player game cash flow of $252,706. Finally, of the 539 players appearing in a 2022-2023 regular season NBA game, we obtain an average salary of $8,274,410. Therefore, a replacement player appearing in all 82 regular season games yields a 2.71% ROI. For complete results, navigate to the public `github` repository at https://github.com/jackson-lautier/nba_roi.

4 Discussion

A vital component of competently investing in capital markets is assessing the ex post financial performance of invested monies. While such assessments are a standard financial calculation generally, difficulties arise when the returns are non-financial, such as on court basketball activities like rebounding, passing, and scoring. This paper attempts to address these challenges by presenting the first known framework to assess the on court performance of NBA players simultaneously within the relative context of salary. Just as the return on a financial investment is relative to the purchase price, a complete evaluation of player performance is enhanced by considering a player's salary. Such calculations are nontrivial, and the interdisciplinary framework we propose is a five-part process that combines theory from statistics, finance, and economics. With the value of NBA franchises reaching billions of US dollars (Wojnarowski, 2022), the need for such tools is now at an all-time high.

Within the five-part ROI framework we propose, the WinLogit is itself a novel approach within the landscape of on court basketball analytics. We take advantage of player tracking data and the relationship of individual and team statistics within a logistic regression model. The result is an informative wealth redistribution tool that is calibrated to replacement player level analysis. Further, it is a per-game model, which yields a new dimension to the field of basketball statistics in the form of break-even calculations for missed games (e.g., Figure 3). Such a calculation is itself timely, as the NBA's governing body has recently implemented strategies to encourage players to avoid missing games (Wimbish, 2023).

The ROI framework we propose in this manuscript and summarize in Figure 1 is intended to be reliable and complete. Nonetheless, the infancy of research into methods to combine on court performance with player salaries in the NBA naturally suggests numerous areas ripe for further study. For example, while not necessary to utilize our ROI framework, we elect to constrain our empirical analysis to a single NBA regular season to ease exposition. Player contracts typically span multiple seasons, and so a more complete empirical analysis would increase the observation period. Further, our empirical estimates do not consider play-off games, which some NBA analysts consider to be a significant component of a player's value (Mahoney, 2019). Hence, the empirical ROI estimates may be updated to include the playoffs. More generally, the calibration of the WinLogit is to wins, whereas other optimization goals are possible (e.g., championships, revenue). Similarly, the SGV model we propose treats games with higher attendance and viewership as more important. An alternative approach might instead prefer to weight games with a significant impact on the standings as more important (though the two are likely correlated). As an example, Özmen (2016) analyzes the marginal contribution of game statistics across various levels of competitiveness in the Euroleague to win probability. Similarly, Teramoto and Cross (2010) is an example of how weighting schemes may differ for playoff games versus regular season games in the NBA. Something similar may be used to model a game's importance.

The models would also benefit from higher precision. This may come through in the form of greater data detail. For example, considering Nielson television ratings, specific ticket prices, or a more refined approach to allocate television revenue. Individual players may get additional credit for off court revenue, such as from jersey sales. A difficulty of these potential enhancements is to obtain detailed data. Higher precision may also be obtained through enhanced calibration. For example, methods exist to refine the quality of a field-goal attempt (e.g., Shortridge et al., 2014; Daly-Grafstein and Bornn, 2019) or account for peer (i.e., teammate) and non-peer effects (e.g., Horrace et al., 2022). Further, the ROI framework overall may benefit from a robustness analysis to swapping out the WinLogit. We do so with Game Score (Sports Reference LLC, 2023b) and Win Score (Berri et al., 2007b), but many other alternatives may be swapped in part II of the ROI framework of Figure 1.

In addition to the statistical aspect, greater precision may be investigated in the financial aspects of the ROI framework of Figure 1 and the derivation of (19). For example, we assume an NBA player's single season salary is paid in one lump sum at time zero. Generally, a player's salary will be paid in installments throughout the regular season. Obtaining more detailed salary payment data will have an impact on the ROI calculations, which may be of interest. Further, we assume all games are played on equally spaced time intervals. This assumption may be explored using financial rate conversion techniques and more precise game dates. Further, an implicit assumption in (19) is that games in the earlier part of the season are given more weight due to the basic conditions of the *time value of money*. Research into the implication of this assumption, such as randomizing the order of the games to calculate a distribution of realized ROI calculations may be prudent. Additionally, the NBA imposes a player salary cap (National Basketball Association, 2018), which includes a team salary floor. Hence, there is an implicit minimum invested, which suggests a type of *risk-free* rate. This may be explored further to offer *Sharpe Ratio* calculations (e.g., Berk and Demarzo, 2007, (11.17)). In addition to the replacement player adjustments employed herein, previous studies such as Niemi (2010) may be helpful for this analysis.

More generally, the WinLogit may also be used in sports injury-related or performance-based studies. For example, Page et al. (2013) look at the effect of minutes played and usage on a player's *production curve* over the course of their career. Within the model, the Game Score (Sports Reference LLC, 2023b) is used as a measure of production. Our WinLogit offers an alternative measure for a similar analysis. Beyond basketball, Theorem 2.1 applies to many sports. Hence, it is of potential interest to replicate this analysis outside of basketball.

A Proofs

Proof of Theorem 2.1. Observe,

$$X_{ij\bullet} - \bar{X}_{ij\bullet} = \sum_{m=1}^{15} X_{ijm} - \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{m=1}^{15} X_{ijm}\right) = \sum_{m=1}^{15} X_{ijm} - 15\bar{X}_{ijm} = \sum_{m=1}^{15}\Big(X_{ijm} - \bar{X}_{ijm}\Big).$$

This proves (6). Next, recall (4) with $\boldsymbol{x}_i^\top = (X_{i1\bullet} - \bar{X}_{i1\bullet}, \ldots, X_{ik\bullet} - \bar{X}_{ik\bullet})^\top$ to write via (6)

$$\begin{aligned}
\text{logit}(p_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} &= \sum_{j=1}^{k} \beta_j (X_{ij\bullet} - \bar{X}_{ij\bullet}) \\
&= \sum_{j=1}^{k} \beta_j \sum_{m=1}^{15} (X_{ijm} - \bar{X}_{ijm}) \\
&= \sum_{m=1}^{15}\sum_{j=1}^{k} \beta_j (X_{ijm} - \bar{X}_{ijm}) = \sum_{m=1}^{15} \boldsymbol{x}_{im}^\top \boldsymbol{\beta} = \sum_{m=1}^{15} \text{logit}(p_{im}).
\end{aligned}$$

□

Proof of Theorem 2.2. For ease of exposition, define $\omega_{gm} := \text{WinLogit}_{gm}$ and $l_{gm} := \text{logit}(p_{gm})$. For standardization, recall (8), (9), and (10) to first write

$$\frac{1}{m^*}\sum_{g=1}^{n/2}\sum_{m\in\mathcal{M}_g} \omega_{gm} = \frac{1}{m^*}\sum_{g=1}^{n/2}\sum_{m\in\mathcal{M}_g}\left(\frac{1}{s(\text{WL})_{m^*}}\left(l_{gm} - \overline{\text{WL}}_{m^*}\right)\frac{1}{\bar{m}} + \frac{1}{\bar{m}}\right)$$

$$
= \frac{1}{m^*} \frac{1}{s(\mathrm{WL})_{m^*}} \left[\frac{1}{m^*} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(l_{gm} - \overline{\mathrm{WL}}_{m^*} \right) \right] + \frac{1}{m^*} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \frac{1}{\bar{m}}
$$

$$
= \frac{1}{\bar{m}}.
$$

Next, ignore the radical to similarly show

$$
\begin{aligned}
\frac{1}{m^*-1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(\omega_{gm} - \frac{1}{\bar{m}} \right)^2 &= \frac{1}{m^*-1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(\frac{1}{s(\mathrm{WL})_{m^*}} \left(l_{gm} - \overline{\mathrm{WL}}_{m^*} \right) \frac{1}{\bar{m}} \right)^2 \\
&= \frac{1}{\bar{m}^2} \frac{1}{s(\mathrm{WL})^2_{m^*}} \frac{1}{m^*-1} \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \left(l_{gm} - \overline{\mathrm{WL}}_{m^*} \right)^2 \\
&= \frac{1}{\bar{m}^2}.
\end{aligned}
$$

For the MLE, it is sufficient to observe $\mathrm{WinLogit}_{gm}(\boldsymbol{\beta})$ is a function of the parameters $\boldsymbol{\beta}$. The result then follows by the invariance property of the MLE (Mukhopadhyay, 2000, Theorem 7.2.1, pg. 250). □

Proof of Theorem 3.1. For ease of exposition, define $\omega^*_{gm} := \mathrm{WinLogit}^*_{gm}$. Observe,

$$
\begin{aligned}
\mathbf{E}\left(\sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \mathrm{SGV}_g \omega^*_{gm} \middle| \omega^*_{gm} \right) &= \sum_{g=1}^{n/2} \mathbf{E}\left(\sum_{m \in \mathcal{M}_g} \mathrm{SGV}_g \omega^*_{gm} \middle| \omega^*_{gm} \right) \\
&= \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \mathbf{E}(\mathrm{SGV}_g \omega^*_{gm} \mid \omega^*_{gm}) \\
&= \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \mathbf{E}(\mathrm{SGV}_g) \omega^*_{gm} \\
&= \mu \sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \mathrm{WinLogit}_{gm}.
\end{aligned}
$$

The proof is then complete by (13). □

References

J. Berk and P. Demarzo (2007). *Corporate Finance, 1st Edition*. Pearson Education, Inc.

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, S. L. Brook, B. Frick, A. J. Fenn and R. Vicente-Mayoral (2005). "The short supply of tall people: Competitive imbalance and the National Basketball Association." *Journal of Economic Issues* **39**, 1029–1041.

D. J. Berri, S. L. Brook and M. B. Schmidt (2007a). "Does one simply need to score to score?" *International Journal of Sport Finance* **2**, 190–205.

D. J. Berri and A. C. Krautmann (2006). "Shirking on the court: Testing for the incentive effects of guaranteed pay." *Economic Inquiry* **44**, 536–546.

D. J. Berri, M. B. Short and S. L. Brook (2007b). *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.

D. Daly-Grafstein and L. Bornn (2019). "Rao-blackwellizing field goal percentage." *Journal of Quantitative Analysis in Sports* **15**, 85–95.

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

K. Goldsberry (2019). *Sprawlball: A Visual Tour of the New Era of the NBA*. HarperCollins.

N. Halevy, E. Y. Chou, A. D. Galinsky and J. K. Murnighan (2012). "When hierarchy wins: Evidence from the National Basketball Association." *Social Psychological and Personality Science* **3**, 398–406.

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

W. C. Horrace, H. Jung and S. Sanders (2022). "Network competition and team chemistry in the NBA." *Journal of Business & Economic Statistics* **40**, 35–49.

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.

S. Keeley (2023). "21 years ago, Herm Edwards reminded us all that 'you play to win the game'." https://www.msn.com/ - Microsoft Network https://www.msn.com/en-us/news/offbeat/21-years-ago-herm-edwards-reminded-us-all-that-you-play-to-win-the-game/ar-AA1j6GWR. Online; accessed 21 December 2023.

S. A. Klugman, H. H. Panjer and G. E. Willmot (2012). *Loss Models: From Data to Decisions, Fourth Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.

J. Kuehn (2017). "Accounting for complementary skill sets: Evaluating individual marginal value to a team in the National Basketball Association." *Economic Inquiry* **55**, 1556–1578.

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.

Z. Lowe (2020). "2020 NBA awards ballot: Zach Lowe's picks for All-NBA, All-Defensive and All-Rookie." https://www.ESPN.com - Entertainment and Sports Programming Network. https://www.espn.com/nba/story/_/id/29541185/2020-nba-awards-ballot-zach-lowe-picks-all-nba-all-defensive-all-rookie. Online; accessed 24 July 2023.

R. Mahoney (2019). "The 16-game player: Why Draymond Green thrives in the NBA playoffs." Sports Illustrated. https://www.si.com/nba/2019/05/29/draymond-green-warriors-raptors-2019-nba-finals-stephen-curry-kevin-durant. Online; accessed 20 July 2023.

N. G. Mankiw (2003). *Principles of Macroeconomics, 3rd Edition*. South-Western College Pub.

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.

D. C. Montgomery (2020). *Design and Analysis of Experiments: Tenth Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.

N. Mukhopadhyay (2000). *Probability and Statistical Inference*. New York, NY: Marcel Dekker.

National Basketball Association (2018). "Highlights of the Collective Bargaining Agreement between the National Basketball Association (NBA) and National Basketball Players Association (NBPA)." nba.com. https://cdn.nba.com/manage/2021/03/2018-19-CBA.pdf. Online; accessed 19 July 2023.

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

J. B. Niemi (2010). "Evaluating individual player contributions to team offense and defense: A model based approach." JSM Proceedings, Section on Statistics in Sports, Vancouver, BC, Canada: American Statistical Association.

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

M. U. Özmen (2016). "Marginal contribution of game statistics to probability of winning at different levels of competition in basketball: Evidence from the Euroleague." *International Journal of Sports Science & Coaching* **11**, 98–107.

G. L. Page, B. J. Barney and A. T. McGuire (2013). "Effect of position, usage rate, and per game minutes played on NBA player production curves." *Journal of Quantitative Analysis in Sports* **9**, 337–345.

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.

S. M. Shea and C. E. Baker (2012). "Calculating Wins over Replacement Player (WORP) for NHL goaltenders." *Journal of Quantitative Analysis in Sports* **8**.

A. Shortridge, K. Goldsberry and M. Adams (2014). "Creating space to shoot: Quantifying spatial relative field goal efficiency in basketball." *Journal of Quantitative Analysis in Sports* **10**, 303–313.

R. Simmons and D. J. Berri (2011). "Mixing the princes and the paupers: Pay and performance in the National Basketball Association." *Labour Economics* **18**, 381–388. The Economics of Sports Labour Markets.

Sports Media Watch (2024). "Major pro and college sports teams ranked by market size." Sports Media Watch. https://www.sportsmediawatch.com/nba-market-size-nfl-mlb-nhl-nielsen-ratings/. Online; accessed 9 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.

Spotrac (2023). "Contract Details." Spotrac.com. https://www.spotrac.com/nba/portland-trail-blazers/chance-comanche-82139/. Online; accessed 25 July 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.

M. Teramoto and C. L. Cross (2010). "Relative importance of performance factors in winning NBA games in regular season versus playoffs." *Journal of Quantitative Analysis in Sports* **6**.

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.

J. Varma (2021). *jrvFinance: Basic finance; NPV/IRR/annuities/bond-Pricing; Black Scholes*. R package version 1.4.3.

J. Wimbish (2023). "NBA CBA: New details emerge about in-season tournament, 65-game threshold for MVP, other awards." https://www.cbssports.com/ - CBS Sports. https://www.cbssports.com/nba/news/nba-cba-new-details-emerge-about-in-season{}-tournament-65-game-threshold-for-mvp-other-awards/. Online; accessed 28 July 2023.

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.

NBA ROI: Supplemental Material

The following is intended as an online companion supplement to the manuscript, *A new framework to estimate return on investment for player salaries in the National Basketball Association*. Please attribute any citations to the original manuscript. This companion includes a detailed literature review, extended details for the WinLogit logistic regression, a robustness analysis for the WinLogit, a simulation study, and an extension to Theorem 3.1. All data and replication code is publicly available at the repository: https://github.com/jackson-lautier/nba_roi.

A Detailed Literature Review

The purpose of this section is to provide more detail to the literature review in the main document. It is to serve as a helpful reference for readers interested in learning more about background material, whereas the main body of the manuscript focuses more on its own results. We proceed in two parts. Section A.1 focuses on basketball performance analysis, especially as it relates to the desired properties of the ROI framework of Figure 1. Section A.2 then focuses on financial performance analysis within basketball and sports more generally.

A.1 WinLogit

Part II of the ROI framework of Figure 1 requires the basketball performance-based calculations to be contained within a single game unit. As summarized in Section 2, a per-game approach offers some advantages.

We now expand on related literature mentioned only briefly in Section 2. Classical regression treatments, such as Berri (1999), do not perform calculations on a game-by-game basis and have become dated considering the advancements in data availability (National Basketball Association, 2023). Data advancements also rule out Page et al. (2007), who fit a hierarchical Bayesian model to 1996-1997 NBA box score data to measure the relative importance of a position to winning basketball games. The same is true for Fearnhead and Taylor (2011), who, in another Bayesian study, propose an NBA player ability assessment model that is calibrated to the relative strength of opponents on the court (via various forms of prior season data; Fearnhead and Taylor (2011) provide results for the 2008-2009 NBA regular season). The work of Casals and Martínez (2013), who fit an OLS model to 2006-2007 NBA regular season data in an attempt to measure the game-to-game variability of a player's contribution to points and Win Score (e.g., Berri et al., 2007b; Berri and Bradbury, 2010), is closer in spirit but does not provide the level of box score detail we desire (the same is true for Martínez (2012)).

A.2 Return on Investment

Parts III, IV, and V of the ROI framework of Figure 1 utilize part II to perform the financial calculations. As we suggest in Section 3, no known studies consider both player salary and on court performance simultaneously.

We now expand on related work mentioned only briefly in Section 3. Idson and Kahane (2000) attempt to derive the determinants of a player's salary in the National Hockey League with a model that incorporates the performance of teammates. We consider the NBA, however, and our methodology differs considerably (see Section 2.1). Berri et al. (2005) identify the importance of height in the NBA and juxtaposes it against population height distributions to explain competitive imbalances observed in the NBA. Such imbalances are thought to negatively impact economic outcomes of sports leagues (Berri et al., 2005). While financial considerations enter into the analysis of Berri et al. (2005), it does not concern the ROI of single players but rather professional leagues overall. Tunaru et al. (2005) develop a claim contingent framework that is connected to an option style valuation of an on field performance index for football players. Our proposed method differs materially, however, and we focus on basketball rather than football.

Berri and Krautmann (2006) find mixed results to the question of whether or not signing a long-term contract leads to shirking behavior from NBA players. The overall objective of their study differs meaningfully from that of our proposed realized ROI metric, however. More recently, Simmons and Berri (2011) find salary inequality is effectively independent of player and team performance in the NBA, a result that runs counter to the hypothesis of fairness in traditional labor economics literature. In a related study, Halevy et al. (2012) find the hierarchical structure of pay in the NBA can enhance performance. Neither study attempts to produce a contractual ROI, however. Kuehn (2017) assumes the ultimate goal of each team is to maximize their expected number of wins to find teammates have a significant impact on an individual player's productivity. Kuehn (2017) subsequently reports that player salaries are determined instead mainly by individual offensive production, which can lead to a misalignment of incentives between individual players and team objectives. Of note, the salary findings of Kuehn (2017) correspond to those of Berri et al. (2007a), a similar study.

B WinLogit: Additional Details

We first present the initial logistic regression results in Section B.1 for reference. Next, in Section B.2, we verify that the WinLogit appropriately captures winning attributes of basketball teams and find it outperforms both GmSc* and WnSc* in this regard.

B.1 Initial Logistic Regression Results

Model selection within statistical analysis can be a complex process (Kutner et al., 2005), often with no clear answer. We detail our approach to decide on the final model presented in Table 1. Nonetheless, in the interest of transparency and reproductive research, we also present the initial model fitting output in Table B1. Such results may provide additional insights or background, which may be used by analysts to deepen understanding of the drivers of winning in the NBA or simply explore alternative models. All data and replication code is publicly available at the repository: https://github.com/jackson-lautier/nba_roi.

Field	Coefficient	Standard Error	Test Statistic	Significance
(Intercept)	-0.015	0.0755	-0.20	
FG2O	0.260	0.0313	8.31	***
FG2X	-0.352	0.0304	-11.58	***
FG3O	0.551	0.0438	12.59	***
FG3X	-0.371	0.0297	-12.51	***
FTMO	0.121	0.0231	5.25	***
FTMX	-0.217	0.0361	-6.01	***
PF	-0.201	0.0231	-8.70	***
AORB	0.377	0.0464	8.11	***
ADRB	0.322	0.0259	12.44	***
STL	0.428	0.0401	10.67	***
BLK	0.128	0.0345	3.70	***
TOV	-0.348	0.0303	-11.49	***
BLKA	-0.002	0.0371	-0.04	
PFD	0.216	0.0333	6.47	***
AST	-0.016	0.0232	-0.68	
SAST	0.072	0.0222	3.24	**
DEFL	0.020	0.0202	0.99	
CHGD	0.513	0.1020	5.03	***
AC2P	0.041	0.0121	3.42	***
C3P	-0.068	0.0143	-4.77	***
OBOX	-0.101	0.0692	-1.46	
DBOX	0.054	0.0247	2.20	*
OLBR	-0.058	0.0487	-1.20	
DLBR	0.023	0.0539	0.42	
DFGO	-0.233	0.0184	-12.67	***
DFGX	0.076	0.0150	5.08	***
DRV	0.001	0.0096	0.08	
ODIS	0.094	0.2062	0.46	
DDIS	-1.104	0.2151	-5.13	***
APM	0.017	0.0036	4.64	***
AST2	0.010	0.0415	0.23	
FAST	0.010	0.0536	0.19	
OCRB	0.305	0.0387	7.87	***
AORC	-0.008	0.0204	-0.37	
DCRB	0.343	0.0350	9.82	***
ADRC	0.024	0.0151	1.59	

Table B1: **Preliminary Logistic Regression**. The initial model fitting as a first step based on team outcomes for the 2022-2023 NBA regular season. Because player tracking data was not available for four games, $n = 2{,}452$. Significant at $\alpha = 0.001$ ($***$), $\alpha = 0.01$ ($**$), and $\alpha = 0.05$ ($*$). Only fields significant at $\alpha = 0.10$ were kept in the final model of Table 1.

B.2 Robustness Analysis

Recall from Section 2.1 that the underlying logistic regression model for the WinLogit is calibrated to wins. Hence, a standard robustness analysis would be to confirm that the WinLogit generates output consistent with this objective. As such, we perform two types of robustness analysis.

The first is to compare the actual team wins of the 2022-2023 NBA regular season against the team total of (10), (11), and (12). In other words, because

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

it is desirable to compare how many wins are allocated to each team by each model with the actual number of wins recorded by each team for the 2022-2023 NBA regular season. We do exactly this in Table B2. Recall $n = 2{,}452$, which implies there are 1,226 wins to be allocated (four games from the 2022-2023 NBA regular season were missing player tracking data). The reported average absolute errors are larger than the now dated 1.67 observed in Berri et al. (2007b, Table 6.8). The standardization tends to pull teams towards the center, and so the larger errors are generally at the very top and bottom of the standings. Of (10), (11), and (12), the WinLogit is the most accurate for both average and median absolute errors by either win total or team rank. One interpretation of these results is that the WinLogit, thanks to its initial calibration to wins, is more attuned to winning than either Game Score or Win Score. On the other hand, the results are comparable, which is impressive given the simplicity of the Game Score and Win Score formulas. Of course, with modern data collection methods and statistical software, the effort necessary to generate the WinLogit estimates is minimal (recall also that all data and replication code is publicly available at the repository: https://github.com/jackson-lautier/nba_roi).

As a second validation, we perform a logistic regression against game outcome using a team's single game total of (10), (11), and (12). We find that both a team's total WinLogit and WnSc* are highly significant to increase team win probability. GmSc* is not significant, though it is likely due to WnSc* and GmSc* being highly correlated. WinLogit registers as the most significant based on a standard variable importance analysis (Kuhn, 2008). This is likely due to the fact that WinLogit uses many more data fields than either GmSc* or WnSc*.

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)
3	PHI	54	45.22 (8.8)	42.81 (11.2)	42.40 (11.6)	5 (2)	7 (4)	6 (3)
4	DEN	53	45.61 (7.4)	44.71 (8.3)	43.52 (9.5)	3 (1)	3 (1)	3 (1)
5	MEM	51	44.44 (6.6)	43.69 (7.3)	42.95 (8.0)	6 (1)	5 (0)	5 (0)
6	CLE	51	42.03 (9.0)	40.89 (10.1)	41.03 (10.0)	10 (4)	18 (12)	18 (12)
7	SAC	48	45.60 (2.4)	44.57 (3.4)	43.89 (4.1)	4 (3)	4 (3)	1 (6)
8	NYK	47	41.19 (5.8)	41.77 (5.2)	41.42 (5.6)	18 (10)	11 (3)	12 (4)
9	BKN	45	42.46 (2.5)	41.31 (3.7)	41.15 (3.8)	9 (0)	13 (4)	16 (7)
10	PHX	45	42.90 (2.1)	41.13 (3.9)	41.12 (3.9)	7 (3)	15 (5)	17 (7)
11	LAC	44	42.03 (2.0)	40.89 (3.1)	40.27 (3.7)	11 (0)	17 (6)	22 (11)
12	MIA	44	36.64 (7.4)	37.89 (6.1)	38.95 (5.1)	27 (15)	26 (14)	25 (13)
13	GSW	43	41.62 (1.4)	42.86 (0.1)	42.29 (0.7)	14 (1)	6 (7)	7 (6)
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)
18	TOR	41	39.23 (1.8)	40.02 (1.0)	41.42 (0.4)	22 (4)	21 (3)	13 (5)
19	OKC	40	40.99 (1.0)	40.75 (0.8)	41.59 (1.6)	19 (0)	19 (0)	11 (8)
20	CHI	39	40.51 (1.5)	41.00 (2.0)	40.52 (1.5)	20 (0)	16 (4)	21 (1)
21	DAL	38	41.36 (3.4)	39.01 (1.0)	39.38 (1.4)	16 (5)	23 (2)	23 (2)
22	UTA	37	41.79 (4.8)	41.68 (4.7)	41.33 (4.3)	13 (9)	12 (10)	15 (7)
23	WAS	35	42.87 (7.9)	41.82 (6.8)	40.92 (5.9)	8 (15)	10 (13)	19 (4)
24	IND	35	38.34 (3.3)	40.28 (5.3)	41.67 (6.7)	24 (0)	20 (4)	10 (14)
25	ORL	34	37.31 (3.3)	38.22 (4.2)	38.60 (4.6)	25 (0)	24 (1)	27 (2)
26	POR	33	36.96 (4.0)	38.21 (5.2)	39.24 (6.2)	26 (0)	25 (1)	24 (2)
27	CHA	27	35.09 (8.1)	37.87 (10.9)	38.83 (11.8)	28 (1)	27 (0)	26 (1)
28	HOU	22	38.59 (16.6)	36.92 (14.9)	37.20 (15.2)	23 (5)	28 (0)	28 (0)
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 B2: **Model Versus Actual Wins**. A comparison of actual versus estimated wins using the winLogit (WL) (10), the Game Score (GS) (11), and the Win Score (WS) (12) models. The absolute errors (ae) are included, and we also report the model rankings (WLR, WSR, GSR) versus the actual team ranking. All results are for the 2022-2023 NBA regular season. The actual wins are adjusted to omit games without player tracking data available (GSW, CHI, MIN, and SAS).

In any subset combination of two, both models each register coefficients as highly significant. In a standard variable importance analysis (Kuhn, 2008), WinLogit always registers as the most important. In a model using only GmSc* and WnSc*, WnSc* registers as the most important. The results of Tables B2 and B3 simultaneously indicate that all three models (10), (11), and (12) have merits, of which WinLogit has the strongest connection to winning (followed by WnSc* and then GmSc*).

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 B3: **Team Level Models and Wins**. A logistic regression using team totals of (10), (11), and (12) against the game outcome for the total sample of 2,452 game outcomes 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. WnSc* and GmSc* are highly correlated, and any subset logistic regression with any combination of two reports each model coefficient as significant at $\alpha = 0.001$ ($***$).

C Simulation Study

We provide a simulation study to verify the results of Theorem 3.1. We estimate WinLogit^*_{gm} for all $g = 1, \ldots, n/2$ and $m = \mathcal{M}_g$, $g = 1, \ldots, n/2$ using data from the 2022-2023 NBA regular season. These estimates correspond to Section 2.2. Thus, $n = 2{,}452$. Further, we assume $\text{SGV}_g \sim \mathcal{N}(\mu = 100, \sigma^2 = 25)$ for all $g = 1, \ldots, 1{,}226$. We run the following simulation for 1,000 replicates. That is, for each replicate, $r = 1, \ldots, 1{,}000$:

1. Simulate 1,226 random variables from a $\mathcal{N}(\mu = 100, \sigma^2 = 25)$ distribution, which we denote by $\widehat{\text{SGV}}_g$, $g = 1, \ldots, 1{,}226$.

2. Compute the product

$$\hat{\mathcal{S}}g = \widehat{\text{SGV}}_g \sum_{m \in \mathcal{M}_g} \text{WinLogit}^*_{gm},$$

   for $g = 1, \ldots, 1{,}226$.

3. Save the result as the summation,

$$\text{Result}_r = \sum_{g=1}^{1{,}226} \hat{\mathcal{S}}g.$$

In doing so, we find an empirical mean of

$$\frac{1}{1{,}000} \sum_{r=1}^{1{,}000} \text{Result}_r = 122{,}605.6,$$

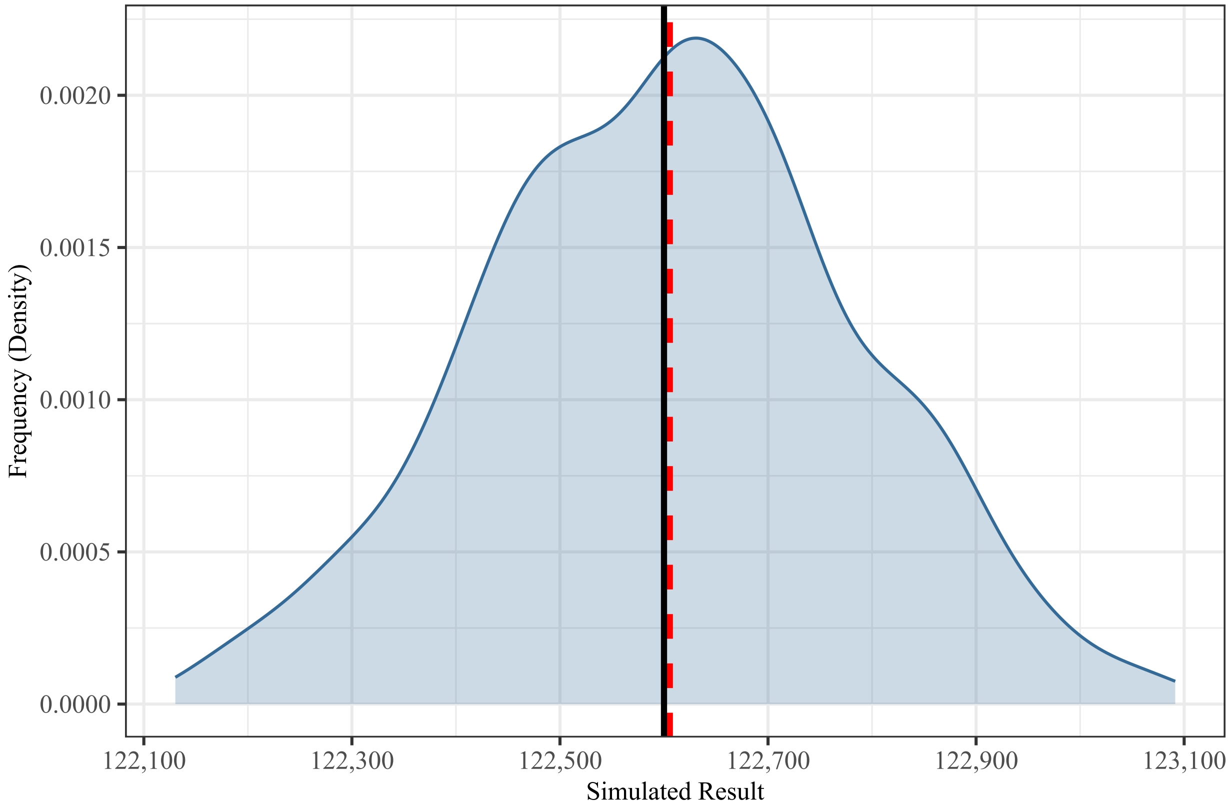

Figure C1: **Simulation Study Results**. A density plot of 1,000 replicates to verify Theorem 3.1. The vertical black line indicates the theoretical mean using Theorem 3.1. The vertical dashed line indicates the empirical sample mean of the 1,000 replicates. The two quantities are quite close, which is a simulation validation of Theorem 3.1.

which is quite close to $\mu(n/2) \equiv 100 \times 1{,}226$. In Figure C1, we provide a density plot of the simulated results.

Finally, we close by stating a minor extension to Theorem 3.1.

Result C.1. *Assume the conditions of Theorem 3.1, and further assume* $\mathrm{Var}(\mathrm{SGV}_g) = \sigma^2$ *for all* $g = 1, \ldots, n/2$. *If* SGV_g *is independent of* SGV_{g^*} *for all* $g, g^* = 1, \ldots, n/2$, $g \neq g^*$, *then*

$$\mathrm{Var}\left(\sum_{g=1}^{n/2} \sum_{m \in \mathcal{M}_g} \mathrm{SGV}_g \mathrm{WinLogit}^*_{gm} \middle| \mathrm{WinLogit}^*_{gm} \right) = \sigma^2 \sum_{g=1}^{n/2} \left(\sum_{m \in \mathcal{M}_g} \mathrm{WinLogit}^*_{gm} \right)^2.$$

Proof. For ease of exposition, define $\omega^*_{gm} := \text{WinLogit}^*_{gm}$. By independence,

$$\begin{aligned}\text{Var}\left(\sum_{g=1}^{n/2}\sum_{m\in\mathcal{M}_g}\text{SGV}_g\omega^*_{gm}\middle|\omega^*_{gm}\right) &= \sum_{g=1}^{n/2}\text{Var}\left(\text{SGV}_g\sum_{m\in\mathcal{M}_g}\omega^*_{gm}\middle|\omega^*_{gm}\right)\\ &= \sum_{g=1}^{n/2}\left(\sum_{m\in\mathcal{M}_g}\omega^*_{gm}\right)^2\text{Var}(\text{SGV}_g)\\ &= \sigma^2\sum_{g=1}^{n/2}\left(\sum_{m\in\mathcal{M}_g}\text{WinLogit}^*_{gm}\right)^2.\end{aligned}$$

□

In an additional simulation study with 10,000 replicates, we obtain an empirical sample variance of the results vector, $\{\text{Result}_r\}_{1\leq r\leq 10{,}000}$, of 32,414.45. This is quite close to the true result, which we calculate to be 31,119.83.

References

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, S. L. Brook, B. Frick, A. J. Fenn and R. Vicente-Mayoral (2005). "The short supply of tall people: Competitive imbalance and the National Basketball Association." *Journal of Economic Issues* **39**, 1029–1041.

D. J. Berri, S. L. Brook and M. B. Schmidt (2007a). "Does one simply need to score to score?" *International Journal of Sport Finance* **2**, 190–205.

D. J. Berri and A. C. Krautmann (2006). "Shirking on the court: Testing for the incentive effects of guaranteed pay." *Economic Inquiry* **44**, 536–546.

D. J. Berri, M. B. Short and S. L. Brook (2007b). *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**.

N. Halevy, E. Y. Chou, A. D. Galinsky and J. K. Murnighan (2012). "When hierarchy wins: Evidence from the National Basketball Association." *Social Psychological and Personality Science* **3**, 398–406.

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.

J. Kuehn (2017). "Accounting for complementary skill sets: Evaluating individual marginal value to a team in the National Basketball Association." *Economic Inquiry* **55**, 1556–1578.

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. 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.

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**.

R. Simmons and D. J. Berri (2011). "Mixing the princes and the paupers: Pay and performance in the National Basketball Association." *Labour Economics* **18**, 381–388. The Economics of Sports Labour Markets.

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.