A new framework to estimate return on investment for player salaries in the National Basketball Association

Jackson P. Lautier*

February 20, 2025

Abstract

An essential component of financial analysis is a comparison of realized returns. These calculations are straightforward when all cash flows have dollar values. Complexities arise if some flows are nonmonetary, 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 flexible five-part procedure that includes a novel player credit estimator, the Wealth Redistribution Merit Share (WRMS). The WRMS is a per-game wealth redistribution estimator that allocates fractional performance-based credit to players standardized and centered to uniformity. We show it is asymptotically unbiased to the natural share and simultaneously more robust. The per-game approach allows for break-even analysis between high-performing players with frequent missed games and average-performing players with consistent availability. The WRMS may be used to allocate 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. We illustrate all methods with empirical estimates from the 2022-2023 NBA regular season. All data and replication code are made available.

Keywords: internal rate of return, load management, player evaluation, player tracking data, sports analytics

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

*Department of Mathematical Sciences, Bentley University, Waltham, MA
Correspondence: jlautier@bentley.edu

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.

For example, to calculate the realized return on investment (ROI) for a sequence of cash flows, it is possible to utilize the *internal rate of return* (IRR) methodology of Berk and Demarzo (2007, §4.8). That is, we solve for the rate of return, r, such that the discounted present value of future return cash flows equals the time zero investment. Formally, let CF_0 be the initial (i.e., negative) investment, and $\mathrm{CF}_1, \ldots, \mathrm{CF}_K$ 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, $t = 1, \ldots K$, are assumed known. Then,

$$\left\{ r : \mathrm{CF}_0 = \sum_{t=1}^{K} \frac{\mathrm{CF}_t}{(1+r)^t} \right\} \tag{1}$$

is the realized ROI. Aside from simple forms of (1), solving for r will typically require the use of optimization software (e.g., Varma, 2021).

Complexities arise when one side of (1) does not have a clear monetary cash value or market price, however. One such case is the player contract in the National Basketball Association (NBA). Specifically, 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. In other words, we endeavor to propose a methodology capable of combining a player's salary and on court performance in such a way as to produce an equivalent formulation of (1). In doing so, we may then solve for r, which is the ROI we desire to estimate.

Financially quantifying on court performance 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 ROI or 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. The first step is to select a measurement period, such as a single NBA regular season. Step two is to select a model to assign fractional credit to players within a single game for all completed games in the measurement time period. Step three is to estimate a Single Game Value (SGV) in dollars for all completed games in the measurement time period. Steps two and three may occur simultaneously after step one. The fourth step is to combine the results of steps two and three to derive player cash flows that are based on relative on court performance. 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 along the lines of (1). The complete ROI process is summarized in Figure 1.

An important component of this proposed framework is the novel player credit estimator we propose, the *Wealth Redistribution Merit Share* (WRMS). It is a general estimator that translates an on court player performance estimate into a standardized fractional share, akin to a wealth redistribution exercise that starts from perfect uniformity and reallocates credit via relative performance. We show the WRMS estimator is asymptotically unbiased to the *natural share*, and it is calibrated to a *replacement player*, often desirable in sports analysis (e.g., Shea and Baker, 2012). The attractiveness of the WRMS is that an analyst is free to choose a player performance estimate, and we present such comparisons. Given we desire to recover (1), our performance measurements are constrained to a single game. This

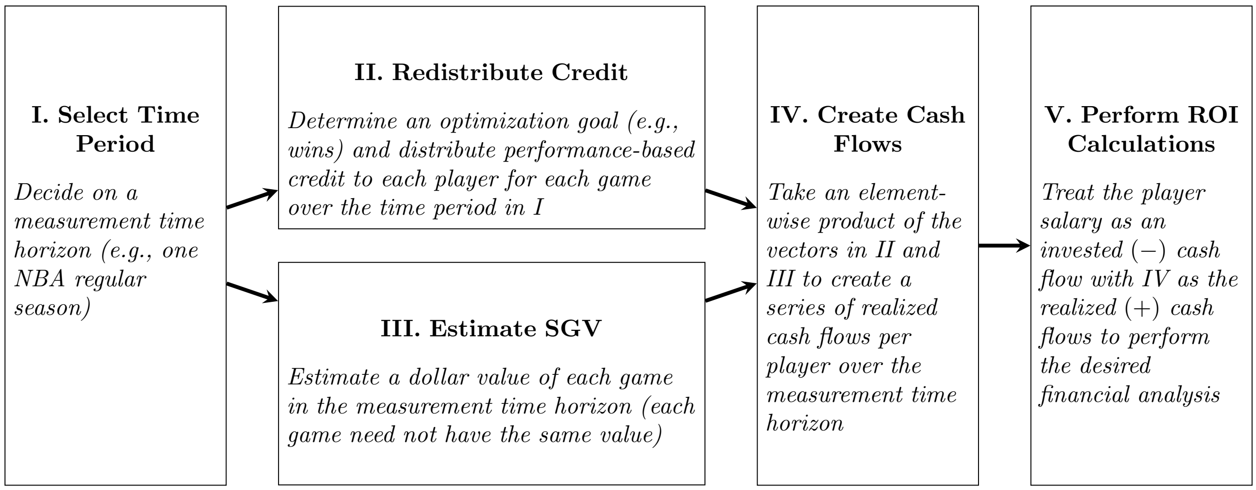

Figure 1: **NBA Contractual ROI Estimation Framework Summary**.

allows us to present a methodology to compare a player with high-performance and frequent missed games against a player with average performance but consistent availability. To our knowledge, such a perspective remains unexplored in the sports analysis literature. We also propose a model based on ticket sales, television revenue, and team standings to estimate the SGV. Conditional on the WRMS estimates, we prove our player share dollar estimates are unbiased to total game value.

The paper proceeds as follows. Section 2 first focuses on measuring player performance. We propose a novel logistic regression player performance model with recent player tracking data, a separate contribution to the applied basketball analysis literature. The details of the ROI framework then occur in Section 3. First, Section 3.1 heuristically derives the WRMS starting from the natural share concept and an assumption of complete naivete. Next, Section 3.2 introduces a model of the SGV, and then Section 3.3 completes the methods necessary to arrive at (1) within an NBA context. Finally, Section 3.4 offers a discussion on how to interpret the ROI results. Throughout Sections 2 and 3, all methods are empirically illustrated with data from the 2022-2023 NBA regular season. The paper concludes in Section 4. The Supplemental Material provides extended details, and all data and replication code used herein may be found at https://github.com/jackson-lautier/nba_roi.

2 Performance Measurement

To evaluate the ROI of player salaries based upon on court performance, it is necessary to evaluate the on court performance of NBA players. This is a topic of significant interest in the field of basketball analytics, and there are many performance measurement models presently in existence (Terner and Franks, 2021). One advantage of the ROI framework of Figure 1 is that there is some freedom in making this choice. On the other hand, there is not a universally accepted standard (e.g., Oliver, 2004; Berri et al., 2007b), and discussion of how to proceed is therefore necessary. As such, the present section will cover four steps.

First, we conduct a literature review of existing methods. This review is briefly summarized herein with extended details available in the Supplemental Material, Section C. Ultimately, we do not find a current performance measurement model that meets our precise needs, and the second step is proposing a novel performance measurement model. We offer a logistic regression model but applied in a novel way that connects team statistics and win probability to individual players. Though applied and ancillary to the main topic of this manuscript, this model is a small contribution itself to the rapidly growing field of basketball analytics. Third, we fit the model using data from the 2022-2023 NBA regular season. The results are summarized herein with vastly extended details in the Supplemental Material, Section G. Finally, we conclude this section with brief comments to reiterate that the ROI results may be applied freely with other choices of performance measurement models.

We begin by reviewing the present literature. Recall we require the basketball performance-based calculations to be contained within a single game unit. This is because we treat 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. Financial interpretations aside, the importance of the single game unit is well-known in the context of basketball analysis (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 both teams use approximately the same number of possessions within one game (Berri et al., 2007b, pg. 101). Finally, our per-game approach to performance measurement implies that running season per game totals allow analysts to determine the exact inflection point of an excellent player that misses many games versus a solid player that consistently plays.

Does an existing performance estimator adequately meet our per-game requirements? Given what is available at present, we believe the answer is largely negative. Many previous studies do not utilize recent player tracking data (e.g., Berri, 1999; Page et al., 2007; Fearnhead and Taylor, 2011; Martínez, 2012; Casals and Martínez, 2013; Martinéz, 2019), and it is our preference to utilize it.[1] 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. For a more detailed literature review, see the Supplemental Material, Section C.

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

$$\begin{aligned} \text{GmSc} = \text{PTS} + 0.4\text{FG} - 0.7\text{FGA} - 0.4(\text{FTA} - \text{FT}) \\ + 0.7\text{ORB} + 0.3\text{DRB} + \text{STL} + 0.7\text{AST} + 0.7\text{BLK} - 0.4\text{PF} - \text{TOV}, \end{aligned} \tag{2}$$

where the abbreviations follow National Basketball Association (2023).[2] Despite the per-game nature of (2), 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

[1] Nonetheless, such models without player tracking data can be effective (e.g., Martinéz (2019) explains over 80% of the variance of home team margin). Both modeling approaches therefore have merit, and we discuss this topic more fully at the conclusion of this section.

[2] A full glossary of common NBA abbreviations may be found in the Supplemental Material, Section D.

potentially unstable over time. Finally, GmSc was derived outside of the peer-review process, which has garnered criticism (e.g., Berri and Bradbury, 2010).

We therefore also consider two additional known on court performance measurement models. The first is the Win Score (WSc) of Berri et al. (2007b), defined as

$$\begin{aligned} \text{WSc} = \text{PTS} + \text{ORB} + \text{DRB} + \text{STL} + 0.5\text{BLK} \\ + 0.5\text{AST} - \text{FGA} - 0.5\text{FTA} - \text{TOV} - 0.5\text{PF}, \end{aligned} \tag{3}$$

which may be instead recoded on a per-game basis.[3] The second is *box plus/minus* (Myers, 2020) (BPM), which may be measured on a per-game basis. BPM is effectively a regression-based performance measurement model that also adjusts for position and team. The calculation is an involved sequence (Myers, 2020) that is difficult to summarize concisely, but the final per-game BPM is readily available (e.g., Sports Reference LLC, 2023a). As with GmSc and WSc, BPM does not use recent NBA data advancements. Further, Myers (2020) recommends caution when using BPM to assess defensive impact.

In addition to these models, we propose a novel, logistic regression-based (Kutner et al., 2005) approach. Suppose there are N total games, indexed by g, $1 \leq g \leq N$. Because each game is either a win or loss, for each g, $1 \leq g \leq N = n/2$, there are two game outcomes, $i = 2g$ and $i = 2g - 1$.[4] Hence, 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, $\mathbf{X}$ (i.e., y_i is a Bernoulli random variable with parameter, p_i, for $1 \leq i \leq n$).

This formulation implies merit performance credit is directly connected to winning games (other performance objectives are also reasonable, see Section 3.4). The binary logistic

[3] A full glossary of common NBA abbreviations may be found in the Supplemental Material, Section D.

[4] As we will introduce another indexing variable, j, for the covariates, we provide an index reference in the Supplemental Material, Section F.

regression model has the form, for $i = 1, \ldots, n$,

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

The form (4) implies

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

Hence, the regression coefficients are called log-odds ratios. That is, $\beta_j \in \boldsymbol{\beta}$, $1 \le j \le k$, 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, $1 \le j, j^* \le k$. 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.

Logistic regression in the context of basketball game outcome data offers some pleasing interpretations. First, if we center each covariate, X_{ij}, i.e., replace X_{ij} with $(X_{ij} - \bar{X}_j)$, where $\bar{X}_j = \sum X_{ij}/n$, then the intercept, β_0, becomes the logit at the mean. In other words, an average game by a team yields a $p(\bar{X}_1, \ldots, \bar{X}_k) = \exp(\beta_0)/(1+\exp(\beta_0))$ 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 (with minor exceptions; e.g., a team turnover is not credited to an individual player). We formalize this 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, $j = 1, \ldots, k$ 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

$$X_{ij\bullet} - \bar{X}_{ij\bullet} = \sum_{m=1}^{15} \Big(X_{ijm} - \bar{X}_{ijm} \Big), \tag{5}$$

where $\bar{X}_{ij\bullet} = \sum_i X_{ij\bullet}/n$ *and* $\bar{X}_{ijm} = \sum_i \sum_m X_{ijm}/15n$*. Further, if we assume* $\beta_0 = 0$*, then*

$$\text{logit}(p_i) = (\boldsymbol{x}_i^*)^\top \boldsymbol{\beta} = \sum_{m=1}^{15} \boldsymbol{x}_{im}^\top \boldsymbol{\beta} = \sum_{m=1}^{15} \text{logit}(p_{im}), \tag{6}$$

where p_i *is the win probability for game outcome* i*,* $i = 1, \ldots, n$*,* $(\boldsymbol{x}_i^*)^\top = (X_{i1\bullet} - \bar{X}_{i1\bullet}, \ldots, X_{ik\bullet} - \bar{X}_{ik\bullet})^\top$*,* $\boldsymbol{x}_{im}^\top = (X_{i1m} - \bar{X}_{i1m}, \ldots, X_{ikm} - \bar{X}_{ikm})^\top$*, and* p_{im} *is the win probability for player* m*,* $m = 1, \ldots, 15$*,*

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

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 the Supplemental Material, Section A. □

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

Remark. *The logistic regression model of* (4) *and Theorem 2.1 assumes each game outcome is independent. Because each game begins at 0-0, and its outcome does not influence another game per NBA rules, this assumption can be reasonably argued. The coefficient estimates of Table 1 are done at the team level, and so the standard errors are reliable. In allocating the team level logit to each player via* (6)*, however, an assumption of independence between players is more dubious. As such, only a point estimate is reported, and there is no attempt to estimate standard errors at the player level. For more discussion, see Section 3.4.*

Remark. *We acknowledge an abuse of notation in the indices appearing in Theorem 2.1. Specifically, when the vector notation appears, we drop the* j *covariate index and shift the player index,* m*, to the* j*th position, e.g.,* (6)*. The player index,* m*, also shifts from game,*

$1 \le m \le 30$, *to team,* $1 \le m \le 15$. *Other equivalent indexing is possible, such as those used in Section 3.1. For an index reference, see the Supplemental Material F.*

We now fit this model to NBA player statistics from the 2022-2023 NBA regular season. To compile an updated set of on court performance 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 36 statistical categories (see the Supplemental Material, Section G for details). For completeness, we note that 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. (See Section 1 for the public repository that contains all data and replication code.)

In fitting the logistic regression and selecting the 36 data fields, we employ three modeling principles: aligning merit to winning, valuing as much on court activity as possible, and avoiding double counting (see the Supplemental Material, Section G.1 for details). The variable selection process consists of first fitting a logistic regression model at the team level for all 36 statistical on court data fields (see the Supplemental Material, Section G.2). We then remove covariates that are not statistically significant at $\alpha = 0.10$ and examine both the resulting significant and non-significant data fields for plausibility within traditional basketball theory (e.g., Oliver, 2004). The complete details of this reasoning may be found in the Supplemental Material, Section G.3. From these selected 20 data fields, we then fit a second logistic regression. In this second model, we estimate $\hat{\beta}_0 = -0.004930$ with a p-value of 0.948. Hence, we may comfortably refit the logistic 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 negligible trade-off (further, the methods of Section 3.1 will correct for this small bias). The final fitted model may be found in Table 1. For reference, the Supplemental Material, Section G contains many additional details of the model fitting process, including an illustrative calculation example for readers less familiar

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: **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).

with the statistical details of the methods employed.

An interpretation of the model in 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 published long-term trend analysis (e.g., Goldsberry, 2019). For an extended interpretation within traditional basketball theory, see the Supplemental Material, Section G.3.

In closing this section, we acknowledge that different analysts may prefer alternative models of on court performance than what we propose in Table 1. For example, alternative models with less input data complexity have been demonstrated to be effective (e.g., Martinéz, 2019). We emphasize that the ROI methods we propose allow an analyst the freedom to make this choice, as long as the evaluation of player performance can be done within a single game. We prefer the model of Table 1 because it uses player tracking data, is designed for a single game evaluation, and, in a robustness analysis, outperforms GmSc, WSc, and BPM (see the Supplemental Material, Section G.4). Nonetheless, we will compare various performance measurement approaches in the forthcoming ROI analysis.

3 Return on Investment

The present section details the various steps necessary to utilize the ROI framework we propose. It is a four-part section that begins by defining the novel WRMS. The objective of the WRMS is to allocate a fractional share of a game's value to each player based upon their on court performance. The WRMS is standardized to a uniform share of wealth and is asymptotically unbiased under mild assumptions. From an economic interpretation, it is a wealth redistribution tool. In the second section, we propose a model for the SGV. It combines the entertainment value and current standings of the two teams involved in a game to estimate a dollar value. In the following section, the results are combined to generate a per game cash flow. This total per game cash flow, as allocated with the WRMS, is unbiased to total expected SGV. The per game cash flows can be combined with a player's salary to calculate the ROI. For reference to the ROI framework of Figure 1, Section 3.1

corresponds to II, Section 3.2 corresponds to III, and Section 3.3 corresponds to IV and V. This section concludes with Section 3.4, which discusses additional areas of further study within the context of interpreting these ROI results. All approaches are illustrated with empirical estimates using data from the 2022-2023 NBA regular season. (See Section 1 for the public repository that contains all data and replication code.)

3.1 Wealth Redistribution Merit Share

We begin by intuitively deriving the WRMS starting from the general concept of a *Natural Share* under an unspecified player performance random variable. Once the WRMS is derived and defined in Theorem 3.1, we then offer brief interpretation comments. Next, we then connect the WRMS to various choices of player performance measurement, such as those introduced and discussed in Section 2. This is done to illustrate the built-in flexibility of the WRMS. We next discuss how the WRMS can be aggregated on a per game cash flow basis, which allows missed games to be treated as defaults. Finally, the WRMS is illustrated with data from the 2022-2023 NBA regular season.

To begin, assume there are $N \geq 1, N \in \mathbb{Z}$ total games over the investment horizon selected in step I of Figure 1. Let the current game be denoted by $g \in \mathbb{Z}$, $1 \leq g \leq N$. Per NBA league rules, we assume each team will roster 15 players (National Basketball Association, 2018), and so 30 players within each game have the potential to contribute. We will index each player by $m \in \mathbb{Z}$, $1 \leq m \leq 30$, for each game, g, $1 \leq g \leq N$. It is desirable to only award players that appear in each game (i.e., $\texttt{MIN} > 0$) with credit.[5] This allows us to treat missed games as *defaults* in the ROI framework. In the sequel, we denote the set of players with positive minutes played in game g, $1 \leq g \leq N$, as $\mathcal{M}_g$, and the set of 30 players with the potential to appear in game g, $1 \leq g \leq N$, as $\overline{\mathcal{M}}_g$. Per NBA rules (National Basketball Association, 2018), a minimum of 10 players (5 per team) will receive playing time (i.e., $\texttt{MIN} > 0$). Formally, then, $10 \leq \#\{\mathcal{M}_g\} \leq \#\{\overline{\mathcal{M}}_g\} = 30$ and $\mathcal{M}_g \subset \overline{\mathcal{M}}_g$.

[5] A full glossary of common NBA abbreviations may be found in the Supplemental Material, Section D.

To calibrate the wealth redistribution estimate based upon on court performance, let us first assume there exists some performance measure, $\Delta_{gm} \in \mathbb{R}$, for each player, m, $m \in \overline{\mathcal{M}}_g$, in each game g, $1 \leq g \leq N$. Hence, the *natural player credit game share*, $\mathcal{N}_{gm}$ for player m, $m \in \overline{\mathcal{M}}_g$, in game g, $1 \leq g \leq N$, becomes

$$\mathcal{N}_{gm} = \frac{\Delta_{gm} \mathbf{1}_{m \in \mathcal{M}_g}}{\sum_{\omega \in \overline{\mathcal{M}}_g} \Delta_{g\omega} \mathbf{1}_{\omega \in \mathcal{M}_g}}, \tag{7}$$

where $\mathbf{1}_q = 1$ if statement q is true and 0 otherwise. It is immediate that $\sum_m \mathcal{N}_{gm} = 1$ for all $1 \leq g \leq N$. Intuitively, this implies that players for both teams compete by way of on court performance for a share of the estimated SGV in dollars. Practically, each player m, $m \in \overline{\mathcal{M}}_g$, for game g, $1 \leq g \leq N$, would receive the $\mathcal{N}_{gm}$ percentage share of the SGV. For any player m, $m \in \{\overline{\mathcal{M}}_g \setminus \mathcal{M}_g\}$, $\mathcal{N}_{gm} = 0$ (i.e., players without playing time receive no credit). All subsequent calculations will build from the natural share construct in (7).

As a starting point, we begin with an assumption of complete naivete. Specifically, we assign a degenerative random variable W for Δ_{gm} such that $\Pr(W = c) = 1$, $c \in \mathbb{R}$, for all m, $m \in \overline{\mathcal{M}}_g$, and g, $1 \leq g \leq N$. In this case, the expected credit share of a player $m \in \mathcal{M}_g$, given the total number of players in the set $\mathcal{M}_g$ is known. It is the uniform share: the inverse of the cardinality of the set $\mathcal{M}_g$. Symbolically, the uniform credit share is $\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g, \Delta_{gm} \sim W) = 1/\#\{\mathcal{M}_g\}$. Hence, we approximate the complete naivete credit share as $1/\mathbf{E}[\#\{\mathcal{M}_g\}]$; that is, the inverse of the average number of players appearing in a game over the measurement time period. If we define $m^* = \sum_g \sum_m \mathbf{1}_{m \in \mathcal{M}_g}$, then an immediate estimator of $1/\mathbf{E}[\#\{\mathcal{M}_g\}]$ is $1/\bar{m}$, where $\bar{m} = m^*/N$. This concept is similar to the starting point of the well-known BPM (Myers, 2020).

To incorporate a version of the *replacement player* standardization widely preferred in sports analysis (e.g., Shea and Baker, 2012), we define the sample statistics

$$\bar{\Delta}_{m^*} = \frac{1}{m^*} \sum_{g=1}^{N} \sum_{m \in \mathcal{M}_g} \Delta_{gm}, \tag{8}$$

and

$$s(\Delta_{m^*}) = \sqrt{\frac{1}{m^*-1}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)^2}. \tag{9}$$

The sample statistics in (8) and (9) provide the WRMS attractive standardization and comparison properties that are robust to the choice of Δ.

Theorem 3.1 (Wealth Redistribution Merit Share)**.** *Assume there are $N \geq 1, N \in \mathbb{Z}$, total games over the investment time horizon. Further assume the set $\mathcal{M}_g$ is known for all g, $1 \leq g \leq N$. Let $\mathcal{S} = \{\Delta_{gm}\}_{1\leq g\leq N, m\in\mathcal{M}_g}$ be a sample of performance measure random variables. Define the wealth redistribution merit share (WRMS) estimator for player m, $m \in \mathcal{M}_g$, for any game g, $1 \leq g \leq N$, as*

$$\mathcal{W}(\mathcal{S})_{gm} = \frac{1}{s(\Delta_{m^*})}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)\frac{1}{\bar{m}}+\frac{1}{\bar{m}}. \tag{10}$$

Then,

(i) The estimator $\mathcal{W}(\mathcal{S})_{gm}$ is standardized to return a sample mean and sample standard deviation of $1/\bar{m}$ for any $\mathcal{S}$. That is,

$$\frac{1}{m^*}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\mathcal{W}(\mathcal{S})_{gm} = \sqrt{\frac{1}{m^*-1}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\mathcal{W}(\mathcal{S})_{gm}-\frac{1}{\bar{m}}\right)^2} = \frac{1}{\bar{m}}.$$

(ii) For any $\mathcal{S}$, $\mathcal{M}_g$ will be known for all g, $1 \leq g \leq N$. If we further assume $\mathcal{S}$ is a sample of identically distributed (though not necessarily independent) performance random variables, then the bias *of $\mathcal{W}(\mathcal{S})_{gm}$ to the conditional natural share, $\mathcal{N}_{gm} \mid \mathcal{M}_g$, denoted by $\text{Bias}(\mathcal{W}(\mathcal{S})_{gm}, \mathcal{N}_{gm} \mid \mathcal{M}_g)$, for all m, $m \in \mathcal{M}_g$, and any g, $1 \leq g \leq N$, is*

$$\text{Bias}(\mathcal{W}(\mathcal{S})_{gm}, \mathcal{N}_{gm} \mid \mathcal{M}_g) = \frac{1}{\bar{m}} - \mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g) = \frac{1}{\bar{m}} - \frac{1}{\#\{\mathcal{M}_g\}},$$

assuming $\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g)$ exists. Further, if $\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g)$ exists, and the number of players appearing in any game g, $1 \leq g \leq N$, is assumed to be an independent and *identically distributed (i.i.d.) random variable with a finite mean, then, as $N \to \infty$,*

$$\text{Bias}(\mathcal{W}(\mathcal{S})_{gm}, \mathcal{N}_{gm} \mid \mathcal{M}_g) \xrightarrow{p} 0.$$

Proof. See the Supplemental Material, Section A. □

Remark. *While we assume $\mathcal{S}$ is identically distributed in statement (ii) of Theorem 3.1, we do not need to assume independence. This allows player performance random variables that may incorporate forms of player dependence to be used, such as Horrace et al. (2022). For conservatism, however, we do not attempt to report confidence intervals or other versions of point estimator uncertainty for the WRMS. We leave this open as an area of further study.*

Remark. *For the asymptotic unbiasedness property of the WRMS, it is assumed the number of players appearing in a game g, $1 \leq g \leq N$ is an i.i.d. random variable with a finite first moment. Per NBA league rules, the number of players appearing in a game will be between 10 and 30 (National Basketball Association, 2018), so the finite first moment will always be satisfied in basketball applications. The i.i.d. assumption can also be reasonably argued, given that each game is a separate entity (see the Remark after Theorem 2.1).*

In an economic interpretation, the WRMS of (10) may be thought of as a prescriptive allocation of the SGV share of wealth earned by a player m, $m \in \mathcal{M}_g$, in reference to the performance measure Δ_{gm}, in comparison to uniformity (i.e., complete naivete) for any game g, $1 \leq g \leq N$. Below average games, (i.e., $\Delta_{gm} < \bar{\Delta}_{m^*}$) will decrease the share below $1/\bar{m}$, and above average games (i.e., $\Delta_{gm} > \bar{\Delta}_{m^*}$) will increase the share above $1/\bar{m}$. In effect, then, (10) is a wealth redistribution tool. That is, starting from the complete naivete assumption that all players appearing in a game have equal performance and thus a perfect uniformity of wealth share, the WRMS then redistributes the wealth to each player based on each player's on court performance in comparison to an average (or replacement) player. A notable property of (10) is that players who perform well on the losing team may still receive a large share of the SGV. Furthermore, the standardization property (i) of Theorem 3.1 will maintain stability in the overall ROI framework while simultaneously providing freedom in the choice of performance measurement. This will facilitate meaningful comparisons, such as those illustrated in Figure 2. Finally, observe that by definition

$$\sum_{g=1}^{N} \sum_{m \in \mathcal{M}_g} \mathcal{W}(S)_{gm} = N, \tag{11}$$

which ensures an unbiased estimate at the aggregate level (i.e., the total reallocation of games sums to the original total of games, N). Because other player share estimates allow for bias (e.g., Berri et al., 2007b; Sports Reference LLC, 2022), we find (11) attractive.

We now connect the performance measurement models of Section 2 to Theorem 3.1. To translate (6) to the performance measurement, Δ_{gm}, $m \in \mathcal{M}_g$, it is necessary to shift the index from game outcome, i, $1 \leq i \leq n$, to game, g, $g = 1, \ldots, n/2$ (recall $N = n/2$). Hence, to use (6) with Theorem 3.1, we obtain the estimator

$$\mathcal{W}(\boldsymbol{X})_{gm} = \frac{1}{s(\mathrm{WL})_{m^*}} \left(\mathrm{logit}(p_{gm}) - \overline{\mathrm{WL}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}, \tag{12}$$

where $\overline{\mathrm{WL}}_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} \mathrm{logit}(p_{gm})/m^*$ and $s(\mathrm{WL})^2_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} (\mathrm{logit}(p_{gm}) - \overline{\mathrm{WL}}_{m^*})^2$ $/(m^* - 1)$. For the sake of performance measurement comparison, we may also use (2) to define the estimator for player m, $m \in \mathcal{M}_g$ in game g, $g = 1, \ldots, n/2$,

$$\mathrm{GmSc}^*_{gm}(\boldsymbol{X}) = \frac{1}{s(\mathrm{GS})_{m^*}} \left(\mathrm{GmSc}_{gm} - \overline{\mathrm{GS}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}, \tag{13}$$

where $\overline{\mathrm{GS}}_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} \mathrm{GmSc}_{gm}/m^*$ and $s(\mathrm{GS})^2_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} (\mathrm{GmSc}_{gm} - \overline{\mathrm{GS}}_{m^*})^2/(m^* - 1)$. Similarly, via (3) we define for player m, $m \in \mathcal{M}_g$ in game g, $g = 1, \ldots, n/2$,

$$\mathrm{WnSc}^*_{gm}(\boldsymbol{X}) = \frac{1}{s(\mathrm{WS})_{m^*}} \left(\mathrm{WnSc}_{gm} - \overline{\mathrm{WS}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}, \tag{14}$$

where $\overline{\text{WS}}_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} \text{WnSc}_{gm}/m^*$ and $s(\text{WS})^2_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} (\text{WnSc}_{gm} - \overline{\text{WS}}_{m^*})^2$ $/(m^* - 1)$. Lastly, for BPM, we may write

$$\text{BPM}^*_{gm}(\boldsymbol{X}) = \frac{1}{s(\text{BPM})_{m^*}} \left(\text{BPM}_{gm} - \overline{\text{BPM}}_{m^*} \right) \frac{1}{\bar{m}} + \frac{1}{\bar{m}}, \tag{15}$$

where $\overline{\text{BPM}}_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} \text{BPM}_{gm}/m^*$ and $s(\text{BPM})^2_{m^*} = \sum_g \sum_{m \in \mathcal{M}_g} (\text{BPM}_{gm} - \overline{\text{BPM}}_{m^*})^2$ $/(m^* - 1)$. By property (i) of Theorem 3.1, all of (13), (14), and (15) remain equivalently standardized to a sample mean and sample standard deviation of $1/\bar{m}$. Hence, we can directly compare wealth allocation differences between (12), (13), (14), and (15) (e.g., Figure 2).

Remark. *It may be tempting to ask why* (7) *cannot be used directly if* $\Delta_{gm} \equiv logit(p_{gm})$ *for all* $m \in \mathcal{M}_g$*, and* g*,* $1 \leq g \leq N$*. The trouble is that, under the assumptions of Theorem 2.1, the conditional natural share in this construct, for any given* m*,* $m \in \mathcal{M}_g$*,* g*,* $1 \leq g \leq N$*, is*

$$\mathcal{N}_{gm} \mid \mathcal{M}_g, \mathbf{X} = \frac{logit(p_{gm})}{\sum_{\omega \in \mathcal{M}_g} logit(p_{g\omega})} \overset{approx}{\sim} \frac{U}{U + V},$$

where $U \sim N(0, \sigma_u^2)$*,* $V \sim N(0, \sigma_v^2)$*, and* $U \perp V$*. This is because, with some abuse of notation and allowance for heuristics,* $logit(p_{gm}) \equiv (\boldsymbol{x}^*_{gm})^\top \boldsymbol{\beta} \overset{approx}{\sim} N(0, \sigma^2)$ *(recall* $\beta_0 = 0$ *by assumption and the covariates are centered). Hence, it can be shown that* $U/(U + V)$ *follows a Cauchy distribution with location parameter* $x_0 = 1/a$ *and scale parameter* $\gamma = \sqrt{a-1}/a$*, where* $a = (\sigma_v^2 + \sigma_u^2)/\sigma_u^2 = \#\{\mathcal{M}_g\}$ *(see the Supplemental Material, Section E). Therefore,* $\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g)$ *does not exist! (The median is the location parameter,* $1/\#\{\mathcal{M}_g\}$*.) Thus, without the stabilization of* (10)*, players would be subject to extreme wealth shares, rendering almost all estimates practically useless. This is an additional advantage of the formulation of* (10) *in that it is robust to the practical use of a logistic regression model for performance measurement, commonly used in the literature (e.g., Teramoto and Cross, 2010; Daly-Grafstein and Bornn, 2019; Terner and Franks, 2021).*

An advantage of the WRMS in (10) is its flexibility. Many choices exist for Δ, such as

(2), (3), or BPM. Different choices for Δ will impact the resulting wealth redistribution, which allows an analyst to tailor player credit by performance measurement preference. To illustrate this, we compare the resulting distributions of (12), (13), (14), and (15) in Figure 2.[6] We see that despite having the same mean and standard deviation of $1/\bar{m} = 4.75\%$, the distributions differ. Specifically, the WRMS and BPM estimates are more symmetric, whereas both the GmSc and WSc are skewed right. In a robustness analysis, we find (12) outperforms all of (13), (14), and (15) in terms of team win prediction and team rank for data from the 2022-2023 NBA regular season (for details, see the Supplemental Material, Section G.4). As such, the remainder of the manuscript will provide results for (12) only, and the Supplemental Material will provide greater discussion on performance measurement comparisons between (12), (13), (14), and (15). This is to focus on the ROI framework we propose rather than the performance measurement details (as Δ is an analyst's choice).

We may also assess the cumulative total performance of a player over the investment period with a financial perspective. Denote $\mathcal{P} = \bigcup_g \overline{\mathcal{M}}_g$ as the set of all players with the potential to contribute over the investment horizon. For a player π, $\pi \in \mathcal{P}$, let $\mathcal{G}_\pi$ represent the set of games for which player π's team appeared (i.e., $\#\{\mathcal{G}_\pi\} = 82$ for a standard NBA regular season). Hence, define for any $g \in \mathcal{G}_\pi$, $\pi \in \mathcal{P}$,

$$\mathcal{W}(\mathcal{S})^*_{g\pi} = \begin{cases} \mathcal{W}(\mathcal{S})_{g\pi}, & \pi \in \mathcal{M}_g \\ 0, & \pi \notin \mathcal{M}_g. \end{cases} \tag{16}$$

Because $\sum_{g=1}^N \sum_{m\in\mathcal{M}_g} \mathcal{W}(\mathcal{S})_{gm} = \sum_{g=1}^N \sum_{\pi\in\overline{\mathcal{M}}_g} \mathcal{W}(\mathcal{S})^*_{g\pi} = N$ still holds trivially, the desirable unbiased property of (11) remains. In financial parlance, the form of (16) implies a

[6] As a data note, the `python` package `basketball_reference_scraper` (Agartha, 2024) was helpful in obtaining game BPM data for the 2022-2023 NBA regular season. (See Section 1 for the public repository that contains all data and replication code.)

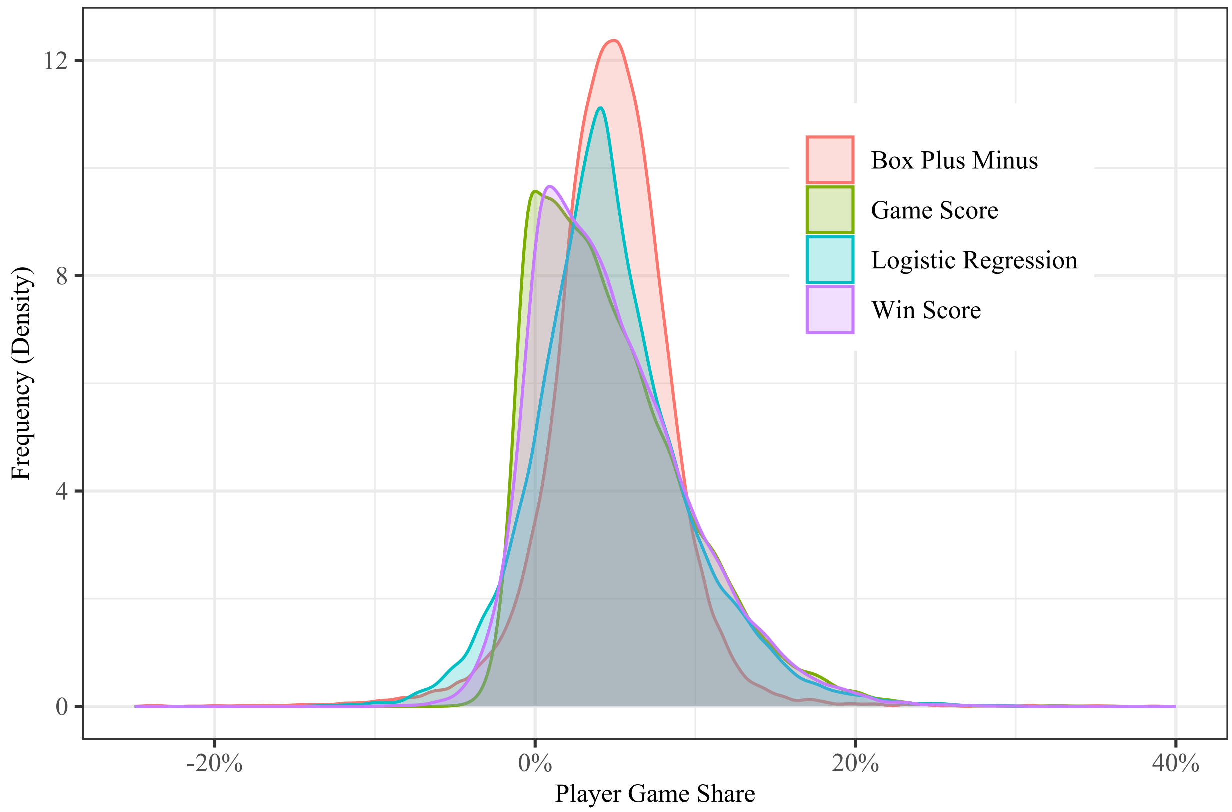

Figure 2: **Wealth Redistribution Comparison**. Standardized frequency distributions of (12), (13), (14), and (15) for all NBA players from the 2022-2023 NBA regular season ($n = 2{,}452$). All four frequency distributions are standardized to have an empirical mean and standard deviation of $1/\bar{m} = 4.75\%$ by Theorem 3.1.

missed game is a *default*. The season total of (16) for player π, $\pi \in \mathcal{P}$, is then

$$\text{PVW}(\cdot)_\pi = \sum_{g \in \mathcal{G}_m} \mathcal{W}(\mathcal{S})^*_{g\pi}. \tag{17}$$

We may consider (17) as a present value of a series of cash flows taking the value of (16) discounted at a zero interest rate. In other words, (17) assumes all SGVs are unity. This allows for a pure performance measure that does not include salary. Notably, the game-by-game approach including zeros used in (16) 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 (17) may offer new insights.

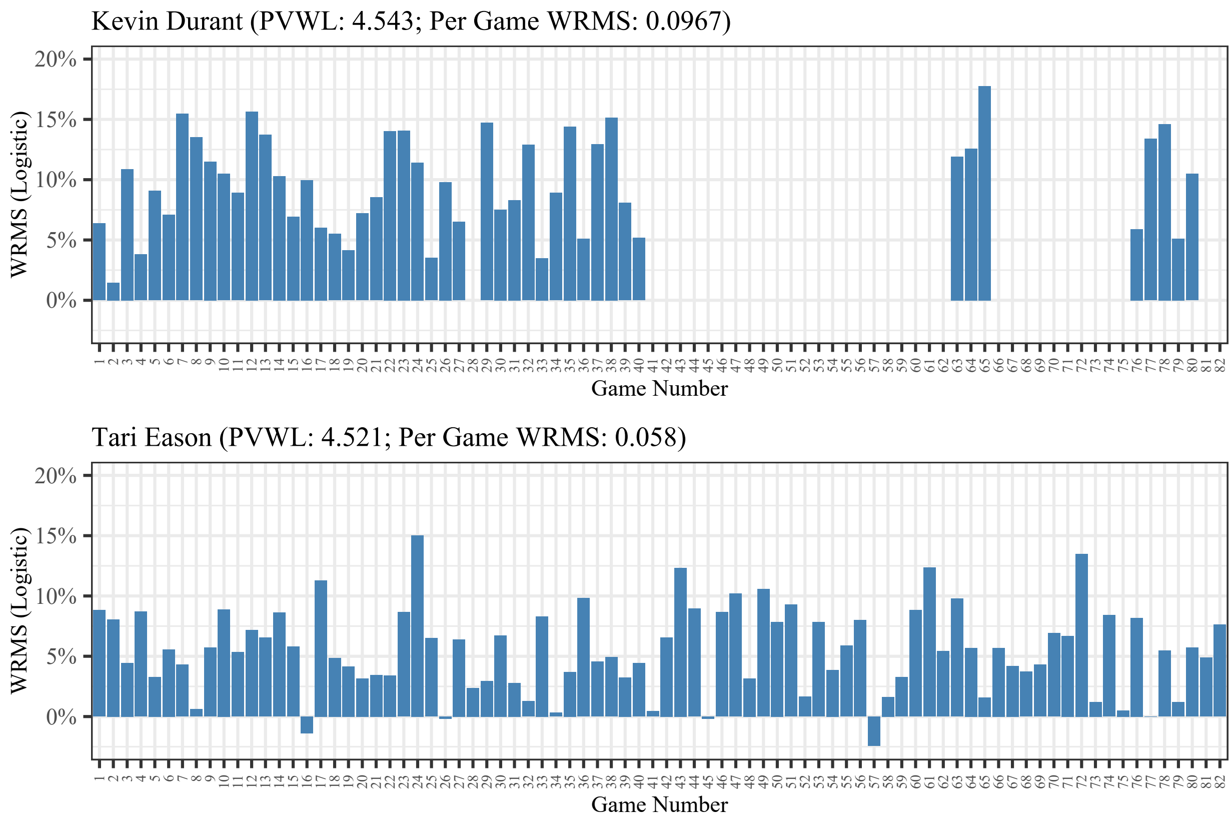

Figure 3: **Quantifying Missed Games**. The per-game approach of (17) allows for break-even calculations between high-performing players with frequent missed games (Kevin Durant, 47 games played, top) against average-performing players with consistent availability (Tari Eason, 82 games played, bottom). Data spans the 2022-2023 NBA regular season.

The placeholder $(\cdot)$ in (17) is general notation to remind us a performance measurement underlies $\mathcal{W}$. We will use PVWL in the sequel to denote (17) that uses (12) for Δ. A summary of the distributions of PVWL by position may be found in Figure 4. We can see the model of Table 1 tends to prefer the center position. In addition, we also report the top performing players, of which Nikola Jokic is the top overall PVWL performer. Though outside the scope of our present analysis, we present a comparison of PVW$(\cdot)$ performance measures using (13), (14), and (15) in the Supplemental Material, Section H. Because $1/\bar{m} = 4.75\%$, an average player playing 82 games would obtain a PV total of 3.896 for the 2022-2023 NBA regular season. This PV total holds regardless of the performance measure used because of the standardization property of the WRMS.

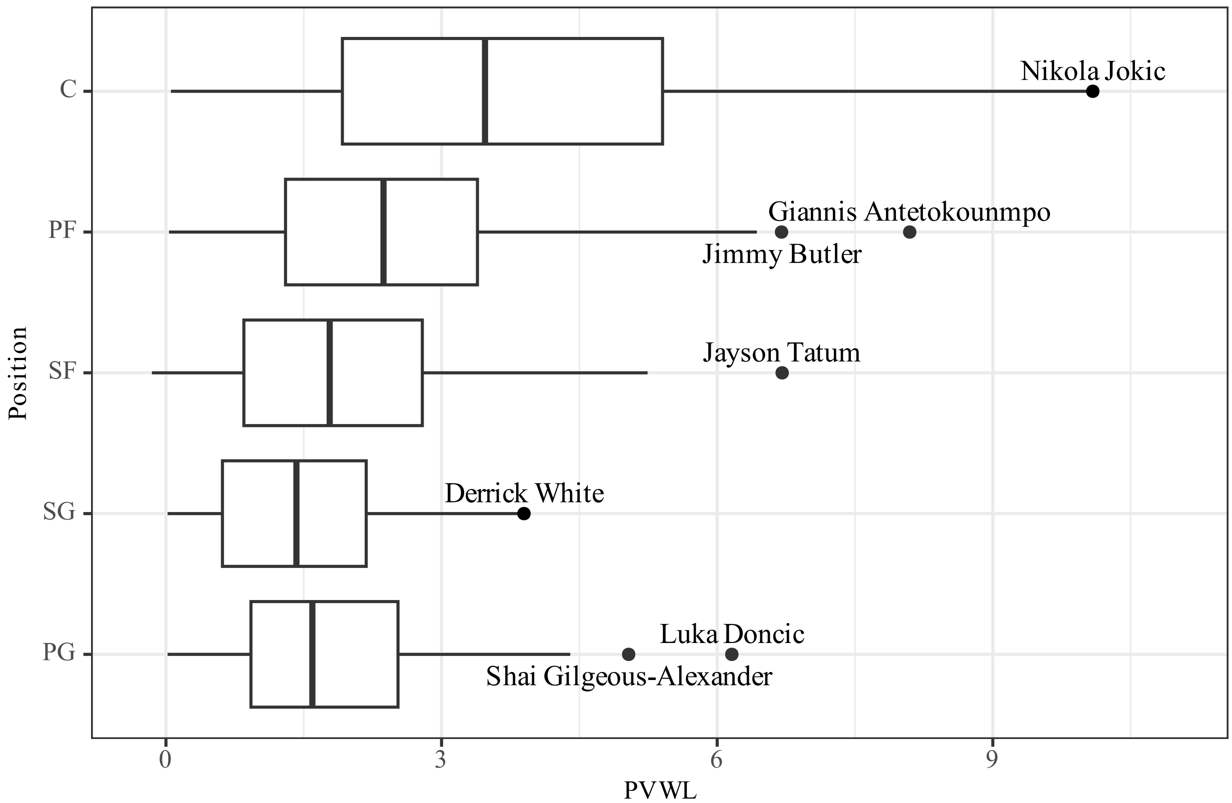

Figure 4: **Top Performers: PVWL**. A summary of the top performers using (17) with logistic regression as the performance measurement (i.e., Table 1) in the WRMS by position. Data spans the 2022-2023 NBA regular season.

3.2 Single Game Value

The methods of the previous section focus on how to redistribute a single unit of game value to each player based upon the player's on court performance. To proceed to the ROI calculation, it is necessary to estimate the SGV in more precise terms. This is the purpose of the present section. It is helpful to imagine each game has a pot of money available, just like a pot in a hand of poker. What would go into this theoretical pot? A first likely component is the entertainment dollar value of the game, which can be estimated through a combination of ticket sales and television revenue. Quite simply, a game that more people want to watch is considered to have more entertainment value than a game with less viewers. A second likely component is the importance of the game to the relative standings of the two teams involved. In other words, a game that takes place between two teams vying for playoff position takes on more importance (i.e., value) than a game between two teams outside of playoff contention. It is our objective, therefore, to propose a SGV model that combines the entertainment value and standings of the two teams involved. We proceed as follows.

For game $g = 1, \ldots, N$, and teams $(j, j') \in \mathcal{T} = \{\texttt{ATL}, \ldots, \texttt{WAS}\}$, $j \neq j'$, define the parametric random variable,

$$\begin{aligned} \mathrm{SGV}_g(\boldsymbol{\alpha}, \boldsymbol{\phi}) = {} & \alpha_1 \mathrm{GATE}_g + \alpha_2 \mathbf{1}_{\mathrm{ESPN}} + \alpha_3 \mathbf{1}_{\mathrm{TNT}} + \alpha_4(\mathbf{1}_{\mathrm{ESPN}} + \mathbf{1}_{\mathrm{TNT}} + \mathbf{1}_{\mathrm{NBATV}}) \\ & + (\phi_j + \phi_{j'})(1 - (\mathbf{1}_{\mathrm{ESPN}} + \mathbf{1}_{\mathrm{TNT}})) + \alpha_5(\mathbf{1}_{\mathrm{TOPSIX}}(j) + \mathbf{1}_{\mathrm{TOPSIX}}(j')), \end{aligned} \tag{18}$$

where the parameter vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)^\top$ 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, α_4, the average advertising revenue for a televised regular season game, and α_5, the estimated single appearance uniform dollar share of total player salary set aside for each regular season game a team plays while in the top six of conference standings; GATE_g is a random variable that represents the attendance for game g, $1 \leq g \leq N$, and ϕ_j is an estimated dollar value of local television revenue for team j, $j \in \mathcal{T}$ (and the same for $j' \in \mathcal{T}$). The indicator functions are defined as $\mathbf{1}_{\mathrm{ESPN}} = 1$ if game g is nationally televised on ESPN and 0 otherwise, $\mathbf{1}_{\mathrm{TNT}} = 1$ if game g is nationally televised on TNT and 0 otherwise, $\mathbf{1}_{\mathrm{NBATV}} = 1$ if game g is nationally televised on NBATV and 0 otherwise, and $\mathbf{1}_{\mathrm{TOPSIX}}(j) = 1$ if team j is in the top-six of conference standings as of the start of game g and 0 otherwise (and the same definition for $\mathbf{1}_{\mathrm{TOPSIX}}(j')$).

Some additional interpretation is warranted. The format of (18) assumes that the estimated dollar value of a regular season game is a sum total of ticket sales, television revenue, advertising revenue, and a playoff pot. We assume each game broadcast on NBATV does not generate television revenue but does generate advertising revenue. In addition, we assume that any game broadcast on national television (i.e., ESPN or TNT) removes local television revenue.[7] For the playoff pot, we assume that all player salaries are ultimately intended to win a championship. Hence, it is assumed the objective of the regular season is to qualify for the playoffs. Though eight teams from each conference qualify for the playoffs, we assume that only the top six teams from each conference have a realistic chance of winning the NBA title (Guerrero, 2022). This also aligns with the recent play-in tournament rule changes in the NBA. Then, in a further alignment with the concept of wealth redistribution, the total sum of all player salaries is evenly distributed to each team for every regular season game appearance as a top six seed. This implicitly assigns more value to games that occur between teams higher in the standings. Greater precision in (18) may be desirable, but there are some data availability constraints. Alternative choices for the SGV model are certainly valid, too, and the ROI framework we propose is flexible enough to allow an analyst to alter (18) as they see fit. For more discussion, see Section 3.4.

We may obtain empirical estimates for the 2022-2023 NBA regular season for the parameter values of (18) using publicly available data sources. Attendance figures are readily available per game (e.g., National Basketball Association, 2023), which allows for a reliable estimate of GATE_g, $g = 1, \ldots, N$. 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, 2023d). 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, $\hat{\alpha}_1 = \$102.64$. 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

[7] Games broadcast on ABC are considered as ESPN, given both networks fall under the Disney umbrella.

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
α_5	Top Six Standings Game Pot	\$4,605,836

Table 2: **$\boldsymbol{\alpha}$ Coefficient Estimates for $\mathbf{SGV}_g$.** Parameter estimates for $\boldsymbol{\alpha}$ for (18) based on available data for the 2022-2023 NBA regular season (National Basketball Association, 2023; Statista, 2023a,d; Lewis, 2023; Statista, 2023c).

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, 2023c). 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 $\hat{\alpha}_4 = \$6{,}080{,}586$.

For the playoff pot, it is first necessary to estimate the total dollar value of NBA salaries. Player salary data for all players from the 2022-2023 NBA regular season are via HoopsHype (2023) (with one supplement for the player Chance Comanche (Spotrac, 2023)). From this data, we estimate total player salary to be \$4,550,565,996. We then identify 988 instances where a team played a regular season game as a top six seed. We thus find $\hat{\alpha}_5 = \$4{,}605{,}836$. A summary of the $\boldsymbol{\alpha}$ coefficient estimates for (18) may be found in Table 2. The estimates for $\boldsymbol{\phi}$ for all 30 teams were compiled by estimating \$1 of revenue for each television home in the local TV market (Nielsen, 2022; Statista, 2023b). For a complete list of all 30 estimates, see the Supplemental Material, Section J. For reference, the top ten teams in terms of total SGV for the 2022-2023 NBA regular season are BKN (\$1,630M), BOS (\$1,629), LAC (\$1,575), PHI (\$1,510), NYK (\$1,504), LAL (\$1,491), PHX (\$1,468), MIL (\$1,412), GSW (\$1,375), and DEN (\$1,362). It is notable that all ten of these teams qualified for the playoffs and many of them consist of the largest television media markets (Sports Media Watch, 2024), each of which helps to validate these estimates. Players on these teams will generate higher ROIs because the games are more valuable per (18), all else equal.

3.3 Return on Investment

With an approach to model the SGVs in hand, we may move to deriving the performance-based cash flows. In doing so, we will have recovered (1), which is the main objective of our analysis. 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, $t = 1, \dots K$, for any player, π, $\pi \in \mathcal{P}$, it is left to multiply (18) with (16) for all $g \in \mathcal{G}_\pi$. This product is player π's, $\pi \in \mathcal{P}$, dollar share of SGV_g, $1 \le g \le N$, based on player π's, $\pi \in \mathcal{P}$, on court performance, and is thus a return cash flow.

Formally, for any player, π, $\pi \in \mathcal{P}$, let $\mathbf{SGV}_{g\in\mathcal{G}_\pi} = (\text{SGV}_1, \dots, \text{SGV}_K)^\top$ be a vector of SGVs, via (18), and let $\mathbf{W}_{g\in\mathcal{G}_\pi} = (\mathcal{W}^*_{1\pi}, \dots, \mathcal{W}^*_{K\pi})^\top$ be a vector of WRMSs, via (16), for all games in which player π's, $\pi \in \mathcal{P}$, team appeared over the investment time horizon, where $\#\{\mathcal{G}_\pi\} = K \in \mathbb{N}$. Then the vector of return cash flows over the investment time horizon for player π, $\pi \in \mathcal{P}$, becomes

$$\mathbf{CF}_\pi = (\mathbf{SGV}_{g\in\mathcal{G}_\pi})^\top \text{diag}(\mathbf{W}_{g\in\mathcal{G}_\pi}) = (\text{SGV}_1\mathcal{W}^*_{1\pi}, \dots, \text{SGV}_K\mathcal{W}^*_{K\pi})^\top, \tag{19}$$

where $\text{diag}(\mathbf{W}_{g\in\mathcal{G}_\pi})$ represents a diagonal $K \times K$ matrix with diagonal $\mathbf{W}_{g\in\mathcal{G}_\pi}$. By the definition of (10), it is possible a particularly bad game may result in $\text{SGV}_t\mathcal{W}^*_{t\pi} < 0$ for some t, $t = 1, \dots, K$ and player π, $\pi \in \mathcal{P}$.

Before proceeding to complete the ROI methodology, we illustrate that the form (19) 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 (19), given a performance model in (10), are unbiased to the expected sum total of all SGVs, i.e., $\mathbf{E}(\sum_g \text{SGV}_g)$. In other words, we do not wish to inadvertently "create" or "eliminate" wealth due to a faulty estimator. This property holds if $\mathbf{E}(\mathrm{SGV}_g) = \mu \in \mathbb{R}$ for all $g = 1, \ldots, N$.

Theorem 3.2. *Let* SGV_g *be a single game value random variable for any game,* $g = 1, \ldots, N$ *such that* $\mathbf{E}(\mathrm{SGV}_g) = \mu \in \mathbb{R}$ *for all* $g = 1, \ldots, N$. *Then, conditional on* $\mathcal{W}^*_{g\pi}$ *for all* π, $\pi \in \mathcal{P}$, $g = 1, \ldots, N$,

$$\mathbf{E}\biggl(\sum_{g=1}^{N} \sum_{\pi \in \overline{\mathcal{M}}_g} \mathrm{SGV}_g \mathcal{W}^*_{g\pi} \bigg| \mathcal{W}^*_{g\pi} \biggr) = \mu N.$$

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

Proof. See the Supplemental Material, Section A. □

Finally, to retrieve the form of (1), let $\boldsymbol{\nu}_\pi = ((1 + r_\pi)^{-1}, \ldots, (1 + r_\pi)^{-K})^\top$ be a vector of discount factors at the rate, r_π, where $\pi \in \mathcal{P}$. Then the contractual ROI for player π, $\pi \in \mathcal{P}$, over the investment time horizon, is the rate, r_π, that equates the discounted present value of player π's, $\pi \in \mathcal{P}$, cash flows, (19), to player π's, $\pi \in \mathcal{P}$, salary. That is,

$$\biggl\{ r_\pi : \mathrm{CF}^\pi_0 = (\mathbf{SGV}_{g \in \mathcal{G}_\pi})^\top \mathrm{diag}(\mathbf{W}_{g \in \mathcal{G}_\pi}) \boldsymbol{\nu}_\pi \equiv \sum_{t=1}^{K} \frac{\mathrm{SGV}_t \mathcal{W}^*_{t\pi}}{(1 + r_\pi)^t} \biggr\}, \tag{20}$$

where CF^π_0 is player π's, $\pi \in \mathcal{P}$, full salary over the investment time horizon. We have thus recovered (1), which completes the ROI framework of Figure 1. We remark that (20) relies on a set of reasonable assumptions, which are discussed more fully in Section 3.4.

We now proceed to provide contractual ROI estimates for NBA players from the 2022-2023 NBA regular season. Though the ROI framework is flexible to the choice of Δ, we will use (12) with the missed game adjustment (16). Our only restriction is that a player's salary is at or above the 2022-2023 league minimum, \$1,017,781 (RealGM, L.L.C., 2024). Because we treat missed games as defaults, the minimum game restriction is just one game played. For the SGV, we use (18) with the empirical estimates from Section 3.2. Results for all players in the 2022-2023 NBA regular season may be found in Figure 5.

The purpose of Figure 5 is to illustrate that the ROI framework we propose is capable

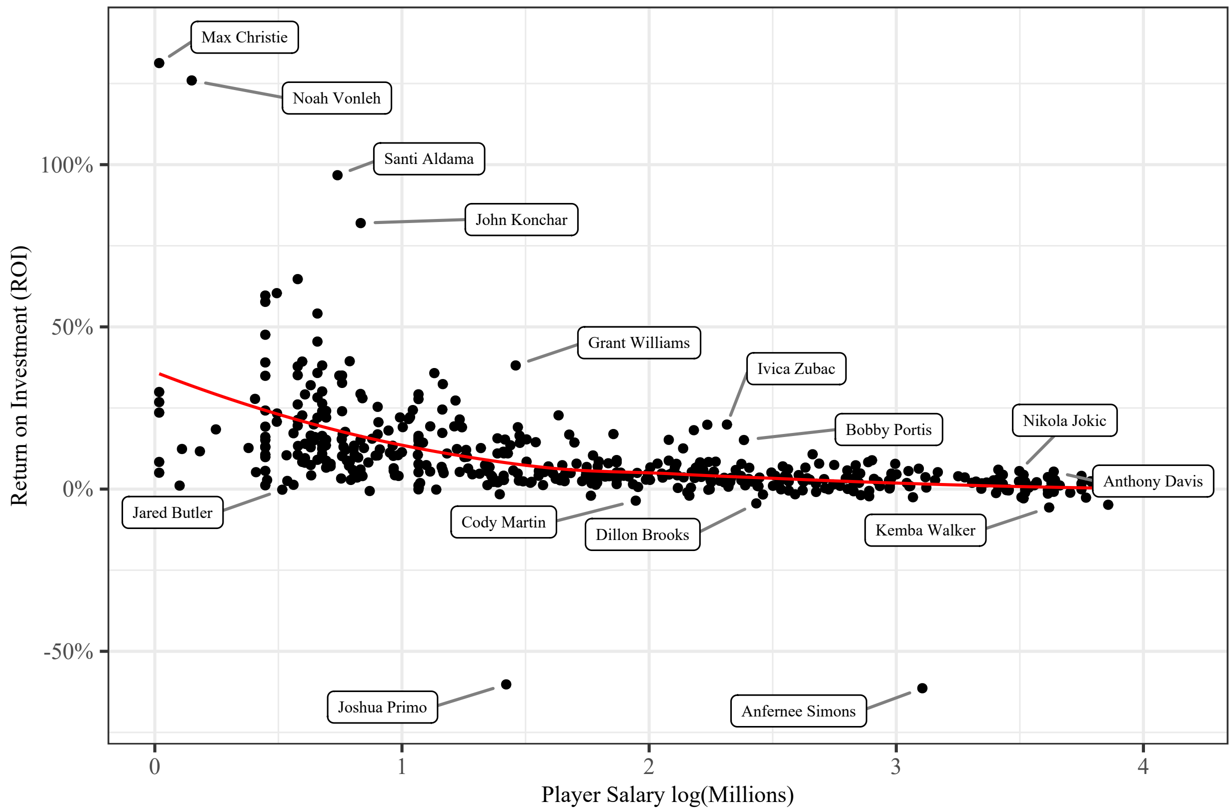

Figure 5: **ROI by Salary: All Players**. A scatter plot of ROI by log of salary for all players (HoopsHype, 2023; Spotrac, 2023) with a salary at the league minimum ($1,017,781 (RealGM, L.L.C., 2024)) or higher for the 2022-2023 NBA regular season. The on court performance measurement is (12) with the missed game adjustment (16). The SGV model follows (18) with empirical estimates from Section 3.2. The ROI calculations may be performed using (20).

of producing the comparative realized ROI of all players in the NBA simultaneously. It is thus a combined measure of player performance that considers on court performance and the player's salary simultaneously. To our knowledge, Figure 5 is the first such attempt to evaluate the contractual ROI for all players in the NBA. Within the ROI framework, it is reasonable that different analysts may prefer one measure of player performance to another or a different model of the SGV. In addition to these considerations, there are other interpretation considerations to Figure 5. These are detailed in Section 3.4.

Nonetheless, Figure 5 illustrates how the ROI framework we propose may be used by NBA teams to target players that may represent a better relative value at various salary ranges. For example, the player performance model of (16) combined with (18) suggests that Grant Williams, Ivica Zubac, and Bobby Portis offered teams an attractive combination of on court performance per salary. On the other hand, the player performance model of (16) combined with (18) suggests that Dillon Brooks, Anfernee Simons, and Kemba Walker did not offer teams an attractive combination of on court performance per salary relative to peers. In this way, different versions of Figure 5 produced using the ROI framework we propose with player performance and SGV models tailored to team preferences may be used to evaluate the performance of NBA team player personnel decision-makers when signing players. Similarly, different versions of Figure 5 may be used by the players or player agents in negotiating a new contract that is more closely aligned with comparable players in the aggregate market. While we believe the the player performance model of (16) combined with (18) is a reasonable approach in that Figure 5 may be used for these purposes directly, both of these models may be changed to meet analyst preferences.

As an additional illustration of the utility of the ROI estimates of Figure 5, we will 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 3. Table 3 suggests that the PF and C positions offer the least risk per unit of return, whereas the SG position is the relative riskiest per unit of return. Such results may be used to help NBA team player personnel decision-makers decide where to invest salary by position, a decision of obvious importance (especially when team salaries are constrained by a *salary cap*, see Section 3.4).

Furthermore, we may calculate a replacement player ROI. Recall we have normalized (10) to $1/\bar{m}$, which is 4.75% for the 2022-2023 NBA regular season. With an average SGV of \$13,598,443, the combination yields a replacement player game cash flow of \$646,089. 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 or average performing player appearing in all 82 regular season games yields a 7.79% ROI. (See Section 1 for the public

Position	Coefficient of Variation
Power Forward (PF)	1.1995
Center (C)	1.2004
Point Guard (PG)	1.2981
Small Forward (SF)	1.4371
Shooting Guard (SG)	2.4309

Table 3: **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 in Figure 5 by position.

repository that contains all data and replication code.)

3.4 Interpretation Considerations

The previous parts of Section 3 have proposed a general framework to estimate the ROI of an NBA player's contract based upon on court performance. Just as in calculating the return of financial assets, however, there are some nuances to how the results may be interpreted. An analogy would be calculating the yield of a bond. The yield follows a formulaic calculation, but the yield itself is either impressive or disappointing depending on its context. In the present section, therefore, we identify potential considerations that extend beyond the ROI framework. It is our hope these considerations will motivate future research.

Extending Calculations Beyond One Year. All illustrative calculations in Section 3.3 consider a single regular season to simplify the presentation of results. Many player contracts are for multiple years, however, and so it is of interest to extend these calculations over the lifespan of a multi-year contract. This may be done by extending (20). The most natural way is to increase K to span multiple seasons (e.g., if two seasons, $K = 164$). The negative salary cash flows could then correspond to the first game of each regular season (e.g., for two seasons, at time 0 and time 83). This is analogous to calculating the yield of a one-year bond versus a multi-year bond. The method is the same, but the amount of cash flows in the equation increases for the multi-year bond. The same is the case for multi-year salaries. Extending beyond one year may also be used to compare different player contract structures, such as a declining annual salary or a back-loaded contract with large annual increases.

Player Contract Objectives. The empirical estimates we present in Section 3.3 do not consider play-off games, which some NBA analysts consider to be a significant component of a player's value (Mahoney, 2019). While the SGVs we present include some consideration for team standings, the empirical ROI estimates may be updated to include playoff games directly. This may be simply including playoff games in (20), or it can be a significant change to how the SGV is measured for playoff games. Even beyond this is how player performance is measured. For example, the logistic regression model in (12) is calibrated to wins, and it may be of interest to explore models calibrated to other performance goals, such as championships or revenue. 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.

Enhanced Statistical Precision. The estimators may benefit from higher precision, such as through 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).

Enhanced Financial Precision. In addition to the statistical aspect, greater precision may be investigated in the purely financial aspects of the ROI framework. 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 (20) 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.

Salary Cap Considerations. Additionally, the NBA restricts how much can be spent on individual player salaries and total team salaries, both maximum and minimums (National Basketball Association, 2018). This is colloquially referred to as a *salary cap* (or *floor*). These restrictions impact any free market assumptions, which can make interpreting an ROI significantly more complex. An analogy would be assessing the performance of an investor under risk restrictions imposed by the mutual fund prospectus. In other words, assessing fund performance should take into consideration that the fund manager can only choose investments that meet the restrictions of the fund prospectus. This can be a complex economic question, however, and we suggest it as an area of further study. One potentially motivating observation is the team salary floor. Hence, there is an implicit minimum invested, which suggests a type of *risk-free* rate. This may be explored further to potentially offer *Sharpe Ratio*-type calculations (e.g., Berk and Demarzo, 2007, (11.17)).

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 in Figure 1, the WRMS of Theorem 3.1 is itself a novel, flexible estimator of player credit capable of considering various estimates of on court player performance. The heuristic derivation of the WRMS suggests a wealth redistribution starting from an assumption of complete naivete. Further, the per-game approach required by (20) 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). Pleasingly, the WRMS is asymptotically unbiased to the natural share. To ensure the ROI framework we propose in this manuscript is reliable and complete, we use a logistic regression model of player performance. The plug and play design of the ROI framework of Figure 1 allows for analysts to swap out player performance measures, estimators of the SGV, or even the WRMS altogether. It is our intention that this flexibility will be viewed as a positive attribute.

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. Suggested areas were detailed in Section 3.4, and they include extending calculations for multiple years, considering other player contract objectives, enhancing both statistical and financial precision, and salary cap considerations. It would also be beneficial to have estimates of the variance of the WRMS rather than point estimates. Depending on the player performance random variable employed, however, such calculations may be complex.

References

V. Agartha (2024). "basketball_reference_scraper: A Python package to access data from Basketball Reference." Python Software Foundation: PyPI. https://pypi.org/project/basketball-reference-scraper/. Online; accessed 21 January 2025.

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*. Stanford 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.

O. J. Guerrero (2022). "5 Lowest seeds ever to win an NBA championship." Sportskeeda. https://www.sportskeeda.com/basketball/5-lowest-seeds-ever-win-nba-championship. Online; accessed 11 February 2025.

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. 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.

J. Martinéz (2019). "A more robust estimation and extension of factors determining production (FDP) of basketball players." **6**, 81–85.

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.

D. Myers (2020). "About Box Plus/Minus (BPM)." Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/about/bpm2.html. Online; accessed 20 January 2025.

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.

Nielsen (2022). "2022-2023 local television market universe estimates." National Media Spots. https://www.nationalmediaspots.com/media-stats/Nielsen-2022-2023-DMA-Ranks.pdf. Online; accessed 11 February 2025.

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, 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. (2024). "2017 CBA Minimum Annual Salary Scale." RealGM.com. https://basketball.realgm.com/nba/info/minimum_scale. Online; accessed 1 May 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 (2022). "NBA Win Shares." Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/about/ws.html. Online; accessed 27 December 2022.

Sports Reference LLC (2023a). "2022-23 NBA Leaders." Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/leagues/NBA_2023_leaders.html. Online; accessed 24 July 2023.

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). "Leading TV markets in Canada in 2023 broadcast year, by number of viewers." https://www.statista.com/ - Statista https://www.statista.com/statistics/791905/leading-tv-markets-canada/. Online; accessed 11 February 2025.

Statista (2023c). "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 (2023d). "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 proofs of all major results, a brief review of discounting cash flows with interest, a detailed literature review, a glossary of standard statistical abbreviations used in the NBA, a result related to generating a Cauchy distribution, a reference of indexing variables, extended logistic regression model details, a comparison of performance measurement models, simulation studies as numeric validation of major results (including an extension to Theorem 3.2), and a listing of local television market sizes by city. Unless otherwise stated, all references are to the main manuscript. All data and replication code is publicly available at the repository: https://github.com/jackson-lautier/nba_roi.

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 (5). 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 (5)

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

□

Proof of Theorem 3.1. For the standardization of (i), recall (8), (9), and (10) to write

$$\begin{aligned}\frac{1}{m^*}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\mathcal{W}(\mathcal{S})_{gm} &= \frac{1}{m^*}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\frac{1}{s(\Delta_{m^*})}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)\frac{1}{\bar{m}}+\frac{1}{\bar{m}}\right)\\ &= \frac{1}{\bar{m}}\frac{1}{s(\Delta_{m^*})}\left[\frac{1}{m^*}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)\right]+\frac{1}{m^*}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\frac{1}{\bar{m}}\\ &= \frac{1}{\bar{m}}.\end{aligned}$$

Next, ignore the radical to similarly show

$$\begin{aligned}\frac{1}{m^*-1}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\mathcal{W}(\mathcal{S})_{gm}-\frac{1}{\bar{m}}\right)^2 &= \frac{1}{m^*-1}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\frac{1}{s(\Delta_{m^*})}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)\frac{1}{\bar{m}}\right)^2\\ &= \frac{1}{\bar{m}^2}\frac{1}{s(\Delta_{m^*})^2}\frac{1}{m^*-1}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)^2\\ &= \frac{1}{\bar{m}^2}.\end{aligned}$$

For (ii), recall (8), (10) and use the assumption that Δ_{gm} are identically distributed random variables for all m, $m \in \mathcal{M}_g$, and g, $1 \le g \le N$, along with the linear property of expectations (Mukhopadhyay, 2000, Theorem 3.3.2, pg. 116) to write

$$\begin{aligned}\mathbf{E}(\mathcal{W}(\mathcal{S})_{gm}-\mathcal{N}_{gm}\mid\mathcal{M}_g) &= \mathbf{E}\left(\frac{1}{s(\Delta_{m^*})}\left(\Delta_{gm}-\bar{\Delta}_{m^*}\right)\frac{1}{\bar{m}}+\frac{1}{\bar{m}}-\mathcal{N}_{gm}\mid\mathcal{M}_g\right)\\ &= \frac{1}{\bar{m}}\left(\mathbf{E}\left(\frac{\Delta_{gm}}{s(\Delta_{m^*})}\right)-\frac{1}{m^*}\sum_{g=1}^{N}\sum_{m\in\mathcal{M}_g}\mathbf{E}\left(\frac{\Delta_{gm}}{s(\Delta_{m^*})}\right)\right)+\frac{1}{\bar{m}}\\ &\quad -\mathbf{E}(\mathcal{N}_{gm}\mid\mathcal{M}_g)\\ &= \frac{1}{\bar{m}}-\mathbf{E}(\mathcal{N}_{gm}\mid\mathcal{M}_g).\end{aligned}$$

Further, given $\mathcal{M}_g$, $m \in \mathcal{M}_g$, and g, $1 \le g \le N$,

$$\mathcal{N}_{gm}\mid\mathcal{M}_g = \frac{\Delta_{gm}}{\sum_{\omega\in\mathcal{M}_g}\Delta_{g\omega}}.$$

But Δ_{gm} are identically distributed random variables for all m, $m \in \mathcal{M}_g$, and g, $1 \leq g \leq N$, and so the distribution of $\mathcal{N}_{gm} \mid \mathcal{M}_g$ is equivalent for all $m \in \mathcal{M}_g$. Thus, assuming $\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g)$ exists,

$$1 = \mathbf{E}\left(\frac{\Delta_{g1} + \ldots + \Delta_{g\#\{\mathcal{M}_g\}}}{\Delta_{g1} + \cdots + \Delta_{g\#\{\mathcal{M}_g\}}}\right) = \sum_{m \in \mathcal{M}_g} \mathbf{E}\left(\frac{\Delta_{gm}}{\Delta_{g1} + \cdots + \Delta_{g\#\{\mathcal{M}_g\}}}\right) = \#\{\mathcal{M}_g\}\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g),$$

for all $m \in \mathcal{M}_g$, and g, $1 \leq g \leq N$. Hence, $\mathbf{E}(\mathcal{N}_{gm} \mid \mathcal{M}_g) = 1/\#\{\mathcal{M}_g\}$. For the asymptotic unbiasedness, recall the number of players appearing in any game, g, $1 \leq g \leq N$, is assumed to be an i.i.d. random variable with a finite first moment. Hence, the expectation is finite and nonzero. Therefore, by the Weak Law of Large Numbers (Lehmann and Casella, 1998, Theorem 8.2, pg. 54-55) and the continuous mapping theorem (Lehmann and Casella, 1998, Corollary 8.11, pg. 58), consistency follows. □

Proof of Theorem 3.2. Observe,

$$\begin{aligned}
\mathbf{E}\left(\sum_{g=1}^{N} \sum_{\pi \in \overline{\mathcal{M}}_g} \mathrm{SGV}_g \mathcal{W}^*_{g\pi} \middle| \mathcal{W}^*_{g\pi}\right) &= \sum_{g=1}^{N} \mathbf{E}\left(\sum_{\pi \in \overline{\mathcal{M}}_g} \mathrm{SGV}_g \mathcal{W}^*_{g\pi} \middle| \mathcal{W}^*_{g\pi}\right) \\
&= \sum_{g=1}^{N} \sum_{\pi \in \overline{\mathcal{M}}_g} \mathbf{E}(\mathrm{SGV}_g \mathcal{W}^*_{g\pi} \mid \mathcal{W}^*_{g\pi}) \\
&= \sum_{g=1}^{N} \sum_{\pi \in \overline{\mathcal{M}}_g} \mathbf{E}(\mathrm{SGV}_g) \mathcal{W}^*_{g\pi} \\
&= \mu \sum_{g=1}^{N} \sum_{m \in \mathcal{M}_g} \mathcal{W}_{gm}.
\end{aligned}$$

The proof is then complete by (11). □

B Financial Review

The objective of the manuscript is to calculate an internal rate of return or realized return on investment for a sequence of cash flows. Such financial parlance may be unfamiliar in

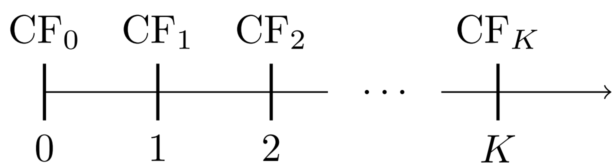

Figure B1: **Cash Flow Time Line**. A classical illustration of a sequence of financial cash flows. The objective of the NBA contractual ROI modeling framework we propose (i.e., Figure 1) is to create a sequence of cash flows in this form, from a combination of salary and on court performance. Once created, it is possible to proceed with standard financial calculations, such as (1).

statistical circles, and we briefly review the fundamentals here. Let us first review *present value*, which relates to the time value of money. For simplicity, suppose we may earn an annual effective rate of i over the next year. Then, if we owe \$1 one year from today, it is sufficient to invest $\$1/(1+i)$ now because

$$\left(\frac{1}{1+i}\right)(1+i) = 1.$$

As such, financial return calculations routinely consider this time value of money. One example is a sequence of cash flows, which is typically represented in a time line, such as Figure B1. In this case, the future cash flows, CF_t, $t = 1, \ldots, K$, represent realized returns. Conversely, the initial time zero cash flow, CF_0, represents the initial investment. To determine the return, we now seek the rate, r such that the initial investment, CF_0, equals the discounted present value of the future cash flows. This is exactly (1) in Section 1. Many references exist with expanded details, such as Berk and Demarzo (2007).

C Detailed Literature Review

The purpose of this section is to provide more detail to the literature review in the main document, which was abbreviated for ease of exposition. We proceed in two parts. Section C.1 focuses on basketball performance analysis, especially as it relates to the desired properties of the ROI framework of Figure 1. Section C.2 then focuses on financial performance analysis within basketball and sports more generally.

C.1 Performance Measurement

Part II of the ROI framework of Figure 1 requires the basketball performance-based calculations to be contained within a single game unit to better mirror financial analysis. As we find in Section 2, a single game performance measurement that also considers more recent player tracking data is not presently available. This motivates the logistic regression analysis we pursue beginning from (4) and expanded upon in Section G. For completeness, we now provide additional detail to the studies referenced 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, 2023b). 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) and Martinéz (2019)).

C.2 Return on Investment

To our knowledge, no basketball studies consider both player salary and on court performance simultaneously. Per the financial aspects of the ROI framework, we expand on related work.

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. 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 ROI framework, 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.

D Basketball Glossary

The main body of the manuscript assumes some familiarity with the NBA, especially the common statistical abbreviations used in the National Basketball Association (2023b). For completeness, we provide a glossary of such abbreviations not defined in the main body of the manuscript (ordered by appearance). All definitions are taken directly from National Basketball Association (2023b), which, for reference, also provides a glossary.

MIN (*Minutes Played*) The number of minutes played by a player or team.

PTS (*Points*) The number of points scored.

FG (*Field Goals Made*) The number of field goals that a player or team has made. This includes both 2 pointers and 3 pointers.

FGA (*Field Goals Attempted*) The number of field goals that a player or team has attempted. This includes both 2 pointers and 3 pointers.

FT (*Free Throws Made*) The number of free throws that a player or team has made.

FTA (*Free Throws Attempted*) The number of free throws that a player or team has made.

ORB (*Offensive Rebounds*) The number of rebounds a player or team has collected while they were on offense.

DRB (*Defensive Rebounds*) The number of rebounds a player or team has collected while they were on defense.

STL (*Steals*) Number of times a defensive player or team takes the ball from a player on offense, causing a turnover.

AST (*Assists*) The number of assists – passes that lead directly to a made basket – by a player.

BLK (*Blocks*) A block occurs when an offensive player attempts a shot, and the defense player tips the ball, blocking their chance to score.

PF (*Personal Fouls*) The number of personal fouls a player or team committed.

TOV (*Turnovers*) A turnover occurs when the player or team on offense loses the ball to the defense.

E Cauchy Distribution

The following result is referenced at the close of Section 2. Suppose $X \sim N(0, \sigma_x^2)$ and $Y \sim N(0, \sigma_y^2)$, where $X \perp Y$. We show

$$\frac{X}{X+Y} \sim \text{Cauchy}\left(x_0 = \frac{\sigma_x^2}{\sigma_y^2 + \sigma_x^2}, \gamma = \frac{\sigma_y \sigma_x}{\sigma_y^2 + \sigma_x^2}\right). \tag{S.1}$$

Recall,

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{y^2}{2\sigma_y^2}\right).$$

Hence, define $Z = X/(X+Y)$ and $W = X$. By the standard Jacobian transformation (e.g., Mukhopadhyay, 2000, Theorem 4.4.1, pg. 192), the joint probability density function of (Z, W) is

$$f_{Z,W}(z,w) = \frac{1}{2\pi}\left|\frac{w}{z^2}\right| \frac{1}{\sigma_x \sigma_y} \exp\left(-\frac{w^2}{b}\right),$$

where

$$b = \left(\frac{1}{\sigma_x^2} + \frac{(1-z)^2}{z^2 \sigma_y^2}\right)^{-1}.$$

The marginal distribution of Z is then

$$\int_{\mathcal{W}} f_{Z,W}(z,w)dw = \frac{1}{\pi} \frac{b}{\sigma_x \sigma_y z^2}.$$

But,

$$\frac{1}{\pi} \frac{b}{\sigma_x \sigma_y z^2} = \left(\pi \frac{\sqrt{a-1}}{a}\left[1 + \left(\frac{z - \frac{1}{a}}{\frac{\sqrt{a-1}}{a}}\right)^2\right]\right)^{-1},$$

where $a = (\sigma_y^2 + \sigma_x^2)/\sigma_x^2$. This is the probability density function of the Cauchy distribution (e.g., Mukhopadhyay, 2000, (1.7.31), pg. 47), which is specified in (S.1). This result may also be confirmed in the simulation studies of Section I.

F A Reference of Indices

The statement of Theorem 2.1 in combination with the WRMS definition of Theorem 3.1 necessitates a series of indexing variables that may be difficult to track. As a reference, we present Table F1. In fitting a logistic regression model to NBA regular season data, there will be n game outcomes. We index each game outcome by i, $1 \leq i \leq n$. Because there are no ties, there will be $n/2 \equiv N$ wins and, similarly, $n/2 \equiv N$ total games. We index each game by g, $1 \leq g \leq N \equiv n/2$, and each game has two game outcomes. Further, we require by Theorem 2.1 that each team roster 15 players for each game. (The roster of 15 is also set by NBA league rules (National Basketball Association, 2018).) This assumption allows us to fit a centered covariate vector at the team level, and then allocate the fitted team level logit to each player depending on the player's individual statistics for game, g, $1 \leq g \leq N \equiv n/2$. Players for each team are indexed by m, $1 \leq m \leq 15$. The covariates are indexed by j, $1 \leq j \leq k$. More generally, players in each game, g, $1 \leq g \leq N$, are indexed by m, $1 \leq m \leq 30$. For clarity, the player index will occasionally switch to ω, such as in the denominator of (7).

To estimate $\mathcal{W}(\boldsymbol{X})$ defined in (12), we shift the calculations away from game outcomes, i, $1 \leq i \leq n$, to games, g, $1 \leq g \leq N \equiv n/2$. This is because we assume all players in a game, g, $1 \leq g \leq N \equiv n/2$, are competing to amass the largest share of game value, as determined by the single-game performance measurement, Δ. By (6), we estimate Δ as the portion of win probability or fitted logit. Finally, Theorem 3.1 restricts the WRMS calculation to the set of players with playing time in a game, g, $1 \leq g \leq N \equiv n/2$. This set is denoted by $\mathcal{M}_g$, $1 \leq g \leq N \equiv n/2$, where $\#\{\mathcal{M}_g\} \leq 30$. When we desire to utilize (16), there is occasion to switch the player index from a basic number index, m, $1 \leq m \leq 30$, to indexing by player name, π, $\pi \in \mathcal{P}$. Note that the sets $\mathcal{M}_g$ and $\overline{\mathcal{M}}_g$ may be equivalently indexed either by m, $1 \leq m \leq 30$, or player name, π, $\pi \in \mathcal{P}$, for any g, $1 \leq g \leq N$.

	Game	Game Outcome	Player			Covariates		
Index	g	i	m			j		
Start	1	1	1			1		
Stop	$N \equiv n/2$	n	15			k		
			1	X_{i11}	$\dots$	X_{ij1}	$\dots$	X_{ik1}
			$\vdots$	$\vdots$	$\ddots$			$\vdots$
		i	m			X_{ijm}		
			$\vdots$	$\vdots$			$\ddots$	$\vdots$
			15	$X_{i1(15)}$	$\dots$		$\dots$	X_{ik15}
	g							
			1	$X_{(i+1)11}$	$\dots$	$X_{(i+1)j1}$	$\dots$	$X_{(i+1)k1}$
			$\vdots$	$\vdots$	$\ddots$			$\vdots$
		$i+1$	m			$X_{(i+1)jm}$		
			$\vdots$	$\vdots$			$\ddots$	$\vdots$
			15	$X_{(i+1)1(15)}$	$\dots$		$\dots$	$X_{(i+1)k15}$

Table F1: **Indexing Levels**. A summary of indexing levels for the WRMS estimator in combination with the logistic regression estimates (i.e., Section G) of performance measurement.

G Logistic Regression Additional Details

The ROI framework proposed in Figure 1 requires a performance measurement random variable or model for Δ. While many examples are possible, we propose an applied logistic regression model for performance measurement that is updated with recent player tracking data. This model is introduced briefly in Section 2, but the details are omitted to allow the manuscript to focus on the larger ROI framework. The present section intends to fill in these omitted details. First, the three modeling principles of aligning merit to winning, valuing as much on court activity as possible, and avoiding double counting will be detailed. Next, the initial model fitting of all 36 data fields will be presented, from which the final model of Table 1 was derived. Third, we provide a discussion of variable selection within the context of basketball theory. Next, we explore a robustness analysis, which finds the logistic regression model in combination with the WRMS outperforms the Win Score, Game Score, and Box Plus/Minus combinations with the WRMS. Finally, the section concludes with a illustration of how to use the logistic regression model to derive player-level estimates.

G.1 Modeling Principles

We employ three principles for data selection and model calibration: aligning merit to winning, valuing as much on court activity as possible, and avoiding double counting. We now discuss each in turn.

Aligning Merit to Winning. We assume that NBA teams are attempting to maximize wins 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, however, such as maximizing championships or maximizing operating income. Concisely, our logistic regression model is calibrated to win probability.

Valuing 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 (2) and (3), 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).

Avoiding 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 non-obvious 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 (2) and (3), 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 logistic regression approach we propose may offer a novel economic perspective that differs from these traditional basketball measures. In addition, these adjustments, in combination with centering each covariate, may help with issues of multicollinearity (Kutner et al., 2005).

G.2 Initial Logistic Regression Results

Our initial covariate space 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 avoid double-counting and minimize multicollinearity concerns. For reference, a glossary of common NBA abbreviations may be found in Section D.

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 in Section 2. Nonetheless, in the interest of transparency and reproductive research, we also present the initial model fitting output in Table G1. 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. For reference, all data and replication code is publicly available at the repository: https://github.com/jackson-lautier/nba_roi.

G.3 Variable Selection

Variable selection in the model of Table 1 was performed using a traditional barrier of statistical significance of each coefficient estimate. While justifiable from a statistical theory point-of-view (Kutner et al., 2005), it is an inexact science. To help give some further credibility to the performance measurement model of Table 1, therefore, we now attempt to justify the inclusion of each term of the 24 terms in Table 1 within the context of basketball theory. We will then provide similar comments on the 12 data fields not ultimately selected. We begin with each of the terms included in Table 1.

To ease exposition, the discussion of included fields will proceed in a bullet list format with fields categorized in groups: scoring, fouling, rebounding, possession, contesting, passing, screening, and moving.

- **Scoring** (FG2O, FG3O, FTMO, FG2X, FG3X, FTMX). Scoring points is fundamental to winning basketball games per NBA rules, and it is not surprising that making two-

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 G1: **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.

point fields goals (FG2O), three-point field goals (FG3O), and free throws (FTMO) all have a positive and significant effect on win probability. Similarly, it is not surprising that missing two-point field goals (FG2X), three-point field goals (FG3X), and free throws (FTMX) all have a negative and significant effect on win probability.

- **Fouling** (PF, PFD, CHGD). Committing fouls (PF) can give an opponent free throws, which are high percentage scoring opportunities. In addition, NBA players are ejected from a game if they commit six personal fouls. Hence, players that commit fouls quickly may be pulled from the game by a coach to avoid "fouling-out" or must alter their approach to be more passive to avoid committing additional fouls. Taken together, it is not surprising to see that PF has a significant and negative effect on win probability, and personal fouls drawn adjusted for charges drawn (PFD) has a significant and positive effect on win probability. Further, drawing a charge (CHGD) has a significant and positive effect on win probability. Drawing a charge is often painful, as an opposing player collides into a stationary defender that has obtained legal guarding position. This can be viewed as "sacrificing one's body" in the interests of winning, which is often inspirational or motivating to the defending team. In addition, it takes a high level of defensive anticipation to draw a charge without committing a blocking foul. Thus, a player that draws more charges may be a sign of a strong defensive player. Finally, a player with an ability to draw charges may be have a positive reputation with officials, which may also contribute to winning.

- **Rebounding** (AORB, ADRB, DBOX, OCRB, DCRB). Rebounding a missed field goal is an essential part of winning basketball games because it either ends an offensive possession for the opponent without a made basket or extends an offensive possession and provides another opportunity to score against an opponent (Oliver, 2004). As such, it is not surprising to see adjusted offensive and adjusted defensive rebounds (i.e., rebounds less contested rebounds) (AORB, ADRB) have a significant and positive effect on win probability. Additionally, the model of Table 1 finds separate value between adjusted defensive rebounds (ADRB) and contested defensive rebounds (DCRB). To consider the difference between the two, it is not unreasonable to suggest that a contested rebound, in which an opponent is within 3.5 feet of the rebounder (National Basketball Association, 2023a), is evidence of obtaining a possession in a moment where possession may go to either team. Such moments may be a sign of a player exhibiting a higher level of effort or intensity, both of which may be motivating for a team. The same holds for offensive rebounds, both adjusted (AORB) and contested (OCRB). Finally, defensive box outs (DBOX) correspond to "the number of times a player made physical contact with an opponent who was actively pursuing a rebound, showed visible progress or strong effort in disadvantaging the opponent, and successfully prevented that opponent from securing the rebound" (National Basketball Association, 2023a). Preventing opponents from securing rebounds would be expected to have a significant and positive effect on win probability, which is what we find in Table 1. Because it is possible to obtain a DBOX hustle statistic without securing a rebound, it is additional information that can be justifiably included in a model of winning players. Furthermore, the box-out technique is fundamental in its importance and is often taught in the earliest stages of youth basketball (Basketball for Coaches, 2024). As such, it is not surprising that all of the coefficients for AORB, ADRB, DBOX, OCRB, and DCRB are positive and significant.

- **Possession** (STL, TOV). The importance of possessing the basketball and the concept of possession is fundamental to basketball analysis (Oliver, 2004). Hence, losing the ball to the opponent via a turnover (TOV) is considered universally negative outcome in the pursuit of winning basketball games (Oliver, 2004). Pleasingly, then, the model of Table 1 finds a significant and negative effect of TOV on win probability. Conversely, taking the ball from the offense via a steal (STL) is the exact inverse, and we find the expected positive and significant effect of STL on win probability.

- **Contesting** (BLK, AC2P, C3P, DFGO, DFGX). From the perspective of a defending team, it is reasonable to suggest there is a different value in contesting an opponent's made basket versus not contesting an opponent's made basket. The same is true if an opponent happens to miss a field goal attempt. The difference lies in the process of the defending team. A contested shot is evidence that the defending team is in proper guarding position and is exerting strong defensive effort. Because of this, it is not surprising that a statistical model finds additional significance of adjusted contested two point shots (AC2P), blocks (BLK), and defended field goals missed (DFGX) above and beyond just missed field goals (FG2X and FG3X). All of these are positive and significant in terms of their effect on win probability. Further, a BLK also may have a motivating effect on the defending team (blocks are often highlights), and they may have a demotivating effect on the offensive team. That is, opponents may be reluctant to shoot near players known for blocking shots, out of fear of ending up on the wrong end of a highlight. Indeed, the four-time defensive player of the year, Rudy Gobert, is known to deter many opponents from even getting close to the basket given his propensity to blocking shots. Hence, a BLK may be viewed as having a different value than a contested shot or even a missed shot by an opponent. Related to this, there is reasonable difference between a made basket and a made defensive contested shot (DFGO). A player that is contesting field goals that are still made baskets by an opponent may indicate a player that is weaker at contesting shots. As such, the player may be unable to materially impact an opponent's shot, even if the player is in strong guarding position, and we see a negative and significant coefficient. The final statistic to discuss in this category is contested three-point shots (C3P). It is at first glance non-intuitive the coefficient is negative and significant. One possible explanation is that a player may be a weaker defender and thus targeted by the opponent, especially in generating high-value three-point field goal attempts. From this point-of-view, a negative and significant coefficient is justifiable.

- **Passing** (APM). Of all the passing categories considered: only adjusted passes made (i.e., passes made less assists, secondary assists, and free throw assists) (APM) was found to have a significant effect on win probability. Its effect was positive, which connects with the classical basketball motivation of "moving the ball" in that teams with offenses that pass the ball more tend to correspond to winning teams. While there are many examples, the 2014 NBA Champion *beautiful game* San Antonio Spurs have been well documented (MacMullan, 2015).

- **Screening** (SAST). Setting a screen or “pick” is a fundamental basketball strategy (i.e., a “pick-and-roll”) of the offensive team as an attempt to create an advantage against the defense (Oliver, 2004). While data on any type of screen is not readily available, there is a quasi-screen category of screen assists (SAST). A SAST is “the number of times an offensive player or team sets a screen for a teammate that directly leads to a made field goal by that teammate” (National Basketball Association, 2023a). As such, it is not surprising the coefficient of SAST is found to positive and significant in terms of its effect on win probability.

- **Moving** (DDIS). The model of Table 1 finds distance traveled by a defensive team (and therefore player) to have a negative and significant effect on win probability. In the context of basketball, it is desirable for the offensive team to get the defense “in rotation”, which involves man-to-man defense devolving into a series of defenders moving away from their primary responsibility and rotating over to cover a now open opponent (i.e., “help-defense”). This often occurs after the offense has obtained an advantageous position, such as requiring a defense to double-team. In addition, a player “lost on defense” is a classical basketball cliché of a player uncertain of his defensive responsibility and is often seen moving too much (i.e., in and out of position). Finally, a weak defensive player may “gamble” on defense, which involves trying to steal a pass with a low probability of success. If unsuccessful, the player will be caught out of position and forced to move more to get back into position (and the defense will also need to rotate, i.e., travel a further distance to compensate).

We now review the statistics not included in the model of Table 1. As done previously, the discussion of omitted fields will proceed in a bullet list format with fields categorized in groups: scoring, fouling, rebounding, possession, contesting, passing, screening, and moving.

- **Scoring**. All scoring categories were found to be significant and thus included.
- **Fouling**. All fouling categories were found to be significant and thus included.
- **Rebounding** (OBOX, AORC, ADRC). As indicated in Table G1, offensive box outs (OBOX), adjusted offensive (defensive) rebound chances (i.e., offensive (defensive) rebound chances less offensive (defensive) rebounds) (AORC, ADRC) were not found to have a significant effect on win probability. Rebound chances occur when "the closest player to the ball at any point in time between when the ball has crossed below the rim to when it is fully rebounded" (National Basketball Association, 2023a). Because we adjust these chances for collected rebounds, it is not unreasonable that proximity to a potential rebound without collecting the rebound would not have significant effect on win probability, positive or negative (it is akin to random chance). Regarding OBOX, it may be supposed that a similar argument can be made as in the previous section for DBOX. A difference, however, is that many teams instruct offensive players to immediately transition to defense once their team takes a shot. This is known as "floor balance", and it is designed to avoid fast break opportunities for the defense after collecting a rebound. Hence, both a positive and negative argument may be made for OBOX, and it is therefore not surprising that the initial model of Table G1 finds no significant effect on win probability.
- **Possession** (OLBR, DLBR, DEFL). It is a classical basketball cliché that players who dive on the floor for loose balls are sacrificing their body for the good of the team (see similar reasoning for CHGD above). As such, it may be surprising that loose balls recovered on offense (OLBR) and defense (DLBR) do not have a significant effect on win probability in the initial model of Table G1. A possible explanation, however, is that a loose ball is itself a random event, typically occurring after a deflection (DEFL), block (BLK), or a dribbling error by the offensive team. As such, many loose balls

are recovered more by random chance than anticipation or effort. Contrast this to rebounding, for example, which always occurs around the basket. Given this, that neither OLBR or DLBR has a significant effect on win probability can be justified in the context of basketball theory. Similarly, a deflection (DEFL) usually occurs by a strong defensive player from a combination of effort, anticipation, and athleticism. Thus, it is expected that DEFL would have a positive and significant effect on win probability. On the other hand, however, a DEFL then typically results in a 50-50 live ball or goes out of bounds, restoring possession to the offense. If a 50-50 ball, then random chance enters again. The defense may gain possession and record a STL or the offense may collect the loose ball and maintain possession. Further, the initial model of Table G1 includes STL, so it is also not surprising that this ambiguous outcome of a DEFL yields a non-significant coefficient.

- **Contesting** (BLKA). From the previous section, it is expected that blocks against (BLKA), i.e., "the number of shots attempted by a player or team that are blocked by a defender" (National Basketball Association, 2023a) would have a negative effect on win probability. On the other hand, there is an argument that a player getting their shot blocked may represent an aggressive offensive player that takes the ball to the rim and is not afraid of challenging the defense. This has a potentially positive effect on winning from a basketball theory point-of-view. Taken together, therefore, it is not unreasonable that the initial model of Table G1 does not find BLKA significant.

- **Passing** (AST, AST2, FAST). The initial model of Table G1 does not find assists (AST), secondary assists (AST2), nor free throw assists (FAST) to have a significant effect on win probability. At first glance, this is surprising. If we observe that all scoring categories are already included in the model and each of AST, AST2, and FAST require a field goal or free throw attempt, then the lack of significance of AST, AST2, and FAST likely highlights instead the importance of not double-counting. In other words, these statistics are obviously important because they indirectly represent points. But, points are already reflected in the model, so there is no additional information (contrast this to APM, which reflects other passes made).

- **Screening**. All screening categories were found to be significant and thus included.
- **Moving** (DRV, ODIS). Both drives (DRV) and offensive distance traveled (ODIS) may have an indeterminate effect on winning from the perspective of basketball theory. On the one hand, driving the ball into the paint and to the rim can be positive, as it puts pressure on the defense. On the other hand, driving the ball into the paint and potentially many defenders without a plan can result in low percentage shots or turnovers. Given this and the number of fields included in the model, no significant effect for DRV is not unreasonable. Regarding ODIS, there is again some ambiguity. For example, a classical "good shooter" may be benefit his team by moving around on offense to shake defenders and get open shots. Conversely, inexperienced offensive players may over-dribble the ball or cut into areas of the floor that actually bring extra defenders towards the ball. Because of this ambiguity, a lack of significant for ODIS in the initial model of Table G1 is not surprising. Contrast this to DDIS, where extra movement from the defense is a sign of the defense breaking down. The inclusion of DDIS and exclusion of ODIS may also indicate the imbalance between offensive and defensive statistics (i.e., there are more offensive statistics, and so ODIS is more noise, whereas the lack of defensive statistics allows DDIS to pick up some information not included in other statistics).

G.4 Robustness Analysis

Recall from Section 2 that the underlying logistic regression model is calibrated to wins. Hence, a standard robustness analysis would be to confirm that the WRMS in combination with the model of Table 1 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 (12), (13), (14), and (15). In other words, because

$$\sum_{g=1}^{N} \sum_{m \in \mathcal{M}_g} \mathcal{W}(\mathcal{S})_{gm} = N,$$

by definition, 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 G2. 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 (12), (13), (14), and (15), the logistic regression 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 logistic regression, thanks to its initial calibration to wins, is more attuned to winning than either Game Score, Win Score, or Box Plus/Minus. On the other hand, the results are comparable, which is notable because Game Score, Win Score, and Box Plus/Minus rely on only traditional box score statistics. Of course, with modern data collection methods and statistical software, the effort necessary to generate the logistic regression estimates is relatively 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 (12), (13), (14), and (15). We find that both a team's total $\mathcal{W}(\boldsymbol{X})$ and WnSc* are highly significant and positively associated with team win probability. GmSc* is not significant, though it is likely due to WnSc* and GmSc* being highly correlated. BPM* is also not significant, though it is not highly correlated with any of the other terms in the model. The most significant is $\mathcal{W}(\boldsymbol{X})$ based on a standard variable importance analysis (Kuhn, 2008). This may be due to the fact that $\mathcal{W}(\boldsymbol{X})$ uses many more data fields than the traditional box score metrics GmSc*, WnSc*, and BPM*. We also test various subset models: $\{\mathcal{W}(\boldsymbol{X}), \text{WnSc}^*\}$, $\{\mathcal{W}(\boldsymbol{X}), \text{GmSc}^*\}$, $\{\mathcal{W}(\boldsymbol{X}), \text{BPM}^*\}$, and $\{\text{WnSc}^*, \text{GmSc}^*, \text{BPM}^*\}$. In doing so, we find BPM* is only significant if $\mathcal{W}(\boldsymbol{X})$ is excluded and GmSc* is significant when either WnSc* or $\mathcal{W}(\boldsymbol{X})$ is excluded. This suggests that all measures have some significant ability to effect win probability, though some measures do not add much when assessed in combination. In a standard variable importance analysis (Kuhn, 2008), $\mathcal{W}(\boldsymbol{X})$ always registers as the most important. In a model using only GmSc* and WnSc*, WnSc* registers as the most important. The results of Tables G2 and G3 simultaneously indicate that all models (12), (13), (14), and (15) have merits, of which $\mathcal{W}(\boldsymbol{X})$ appears to have the strongest connection to winning as assessed by effect on win probability (followed by WnSc*, then GmSc*, and finally BPM*).

G.5 Illustrative Example

The purpose of the present section is to illustrate the steps needed to perform the logistic regression proposed in Section 2. It is designed to benefit readers with less of a background in statistics interested in implementing the logistic regression performance measurement model we propose. As such, we proceed using a numerical example constructed from the October 22, 2022 NBA regular season game between the Philadelphia 76ers (PHI) and the Boston Celtics (BOS). For ease of exposition, the fields will connect directly with those proposed in Table 1 (though `PTS` are not used in the model). The counting statistics may be found in Table G4. It is our intention to illustrate how a user takes the inputs of Table G4 and fits the logistic regression model of Table 1. These are calculations found in Section 2. Next, we show how the fitted logistic regression model then translates into player level performance measurement estimates, denoted $\text{logit}(p_{gm})$ for player m, $1 \leq m \leq 15$, in game g, $1 \leq g \leq N$. Finally, we make the connection between the player-level logit and team-level logit. In effect,

Rank	Team	Wins	WL (ae)	WS (ae)	GS (ae)	BPM (ae)	WLR (ae)	WSR (ae)	GSR (ae)	BPMR (ae)
		Median Error	3.66	4.95	4.82	5.25	1.00	3.00	4.00	5.00
		Average Error	5.49	5.99	6.47	6.35	2.87	3.93	4.87	5.27
1	MIL	58	46.08 (11.9)	45.08 (12.9)	42.13 (15.9)	43.21 (14.8)	1 (0)	2 (1)	9 (8)	4 (3)
2	BOS	57	45.78 (11.2)	45.60 (11.4)	43.71 (13.3)	44.85 (12.1)	2 (0)	1 (1)	2 (0)	1 (1)
3	PHI	54	45.22 (8.8)	42.81 (11.2)	42.40 (11.6)	43.05 (11.0)	5 (2)	7 (4)	6 (3)	7 (4)
4	DEN	53	45.61 (7.4)	44.71 (8.3)	43.52 (9.5)	42.21 (10.8)	3 (1)	3 (1)	3 (1)	13 (9)
5	MEM	51	44.44 (6.6)	43.69 (7.3)	42.95 (8.0)	44.65 (6.4)	6 (1)	5 (0)	5 (0)	2 (3)
6	CLE	51	42.03 (9.0)	40.89 (10.1)	41.03 (10.0)	42.36 (8.6)	10 (4)	18 (12)	18 (12)	11 (5)
7	SAC	48	45.60 (2.4)	44.57 (3.4)	43.89 (4.1)	43.15 (4.8)	4 (3)	4 (3)	1 (6)	5 (2)
8	NYK	47	41.19 (5.8)	41.77 (5.2)	41.42 (5.6)	41.32 (5.7)	18 (10)	11 (3)	12 (4)	17 (9)
9	BKN	45	42.46 (2.5)	41.31 (3.7)	41.15 (3.8)	39.35 (5.7)	9 (0)	13 (4)	16 (7)	23 (14)
10	PHX	45	42.90 (2.1)	41.13 (3.9)	41.12 (3.9)	44.15 (0.8)	7 (3)	15 (5)	17 (7)	3 (7)
11	LAC	44	42.03 (2.0)	40.89 (3.1)	40.27 (3.7)	42.57 (1.4)	11 (0)	17 (6)	22 (11)	10 (1)
12	MIA	44	36.64 (7.4)	37.89 (6.1)	38.95 (5.1)	36.77 (7.2)	27 (15)	26 (14)	25 (13)	27 (15)
13	GSW	43	41.62 (1.4)	42.86 (0.1)	42.29 (0.7)	40.47 (2.5)	14 (1)	6 (7)	7 (6)	19 (6)
14	LAL	43	41.96 (1.0)	42.74 (0.3)	42.22 (0.8)	40.46 (2.5)	12 (2)	8 (6)	8 (6)	20 (6)
15	NOP	42	41.56 (0.4)	41.27 (0.7)	41.40 (0.6)	43.06 (1.1)	15 (0)	14 (1)	14 (1)	6 (9)
16	ATL	41	41.24 (0.2)	42.69 (1.7)	43.10 (2.1)	42.58 (1.6)	17 (1)	9 (7)	4 (12)	9 (7)
17	MIN	41	40.26 (0.7)	40.00 (1.0)	40.54 (0.5)	42.98 (2.0)	21 (4)	22 (5)	20 (3)	8 (9)
18	TOR	41	39.23 (1.8)	40.02 (1.0)	41.42 (0.4)	40.12 (0.9)	22 (4)	21 (3)	13 (5)	22 (4)
19	OKC	40	40.99 (1.0)	40.75 (0.8)	41.59 (1.6)	42.09 (2.1)	19 (0)	19 (0)	11 (8)	14 (5)
20	CHI	39	40.51 (1.5)	41.00 (2.0)	40.52 (1.5)	42.36 (3.4)	20 (0)	16 (4)	21 (1)	12 (8)
21	DAL	38	41.36 (3.4)	39.01 (1.0)	39.38 (1.4)	40.13 (2.1)	16 (5)	23 (2)	23 (2)	21 (0)
22	UTA	37	41.79 (4.8)	41.68 (4.7)	41.33 (4.3)	41.56 (4.6)	13 (9)	12 (10)	15 (7)	16 (6)
23	WAS	35	42.87 (7.9)	41.82 (6.8)	40.92 (5.9)	41.92 (6.9)	8 (15)	10 (13)	19 (4)	15 (8)
24	IND	35	38.34 (3.3)	40.28 (5.3)	41.67 (6.7)	38.40 (3.4)	24 (0)	20 (4)	10 (14)	24 (0)
25	ORL	34	37.31 (3.3)	38.22 (4.2)	38.60 (4.6)	40.71 (6.7)	25 (0)	24 (1)	27 (2)	18 (7)
26	POR	33	36.96 (4.0)	38.21 (5.2)	39.24 (6.2)	35.54 (2.5)	26 (0)	25 (1)	24 (2)	29 (3)
27	CHA	27	35.09 (8.1)	37.87 (10.9)	38.83 (11.8)	36.78 (9.8)	28 (1)	27 (0)	26 (1)	26 (1)
28	HOU	22	38.59 (16.6)	36.92 (14.9)	37.20 (15.2)	38.34 (16.3)	23 (5)	28 (0)	28 (0)	25 (3)
29	SAS	21	33.67 (12.7)	35.96 (15.0)	37.05 (16.1)	35.29 (14.3)	29 (0)	29 (0)	29 (0)	30 (1)
30	DET	17	32.68 (15.7)	34.37 (17.4)	36.18 (19.2)	35.57 (18.6)	30 (0)	30 (0)	30 (0)	28 (2)

Table G2: **Model Versus Actual Wins**. A comparison of actual versus estimated wins using the $\mathcal{W}(\boldsymbol{X})$ (WL) (12), the Game Score (GS) (13), the Win Score (WS) (14), and the Box Plus/Minus (15) models. The absolute errors (ae) are included, and we also report the model rankings (WLR, WSR, GSR, BPMR) 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).

Field	Coefficient	Standard Error	Test Statistic	Significance
(Intercept)	-14.302	0.6493	-22.03	$***$
$\mathcal{W}(\boldsymbol{X})$	17.770	1.2218	14.54	$***$
WnSc*	10.523	2.5419	4.14	$***$
GmSc*	0.881	2.2567	0.39	
BPM*	0.071	0.4357	0.16	

Table G3: **Team Level Models and Wins**. A logistic regression using team totals of (12), (13), (14), and (15) 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. Other subset logistic regression models tested: $\{\mathcal{W}(\boldsymbol{X}), \text{WnSc}^*\}$, $\{\mathcal{W}(\boldsymbol{X}), \text{GmSc}^*\}$, $\{\mathcal{W}(\boldsymbol{X}), \text{BPM}^*\}$, and $\{\text{WnSc}^*, \text{GmSc}^*, \text{BPM}^*\}$ suggest BPM* is only significant if $\mathcal{W}(\boldsymbol{X})$ is excluded and GmSc* is significant when either WnSc* or $\mathcal{W}(\boldsymbol{X})$ is excluded.

this final illustration is a numerical demonstration of Theorem 2.1. These calculations may also be replicated by following the replication code posted at https://github.com/jackson-lautier/nba_roi.

Consider first fitting the logistic regression model. The first step is to obtain the team level statistics represented by the two "Team Totals" rows of Table G4. Hence, the game in Table G4 represents two inputs, a row with an outcome of 0 (i.e., loss) for PHI and a row with an outcome of 1 (i.e., win) for BOS. In the model summarized in Table 1, there are $n = 2{,}452$ such observations used to estimate the coefficients. Prior to finding the coefficient estimates in Table 1, it is necessary to center each column of regression data. The centering occurs across all $n = 2{,}452$ observations. Hence, the coefficients of Table 1 are found using the "Team Total" rows of Table G5, with the same outcome of 0 for PHI and 1 for BOS. By forcing the intercept to be zero, this centering ensures that an exactly average game by any team results in a fitted win probability of 0.50 (i.e., $\exp(0)/(1+\exp(0))$).

To review some calculations, let $\boldsymbol{X}_{\text{PHI}}$ represent the "Team Totals" row of Table G5 for PHI. Further, let $\hat{\boldsymbol{\beta}}$ represent the Coefficient Estimate column of Table 1. Then,

$$(\boldsymbol{X}_{\text{PHI}})^\top \hat{\boldsymbol{\beta}} = -2.34.$$

This corresponds to a win probability of $\exp(-2.34)/(1+\exp(-2.34)) = 0.0879$ for PHI. Similarly, if $\boldsymbol{X}_{\text{BOS}}$ represents the "Team Totals" row of Table G5 for BOS, then

$$(\boldsymbol{X}_{\text{BOS}})^\top \hat{\boldsymbol{\beta}} = 4.21,$$

which corresponds to a win probability of 0.9854.

The next step is find the player level logits. This returns us to Table G5. Notice first that both PHI and BOS have five "Active Player" rows added when compared with Table G4. This is done to use Theorem 2.1, which requires that each team have the same number of players. All "Active Player" rows contain zeros in all data point entries for the counting statistics of Table G4. Each player is then centered across all entries for all games. With the required 15 players for each game and $n = 2{,}452$, the total centered player data contains 36,780 rows. As an illustration, suppose $\boldsymbol{x}_{\text{HARRIS}}$ corresponds to the Tobias Harris row of Table G5 (excluding the final two columns). Then,

$$(\boldsymbol{x}_{\text{HARRIS}})^\top \hat{\boldsymbol{\beta}} = 0.59.$$

This calculation may be repeated for all 15 players for each team, and it may be verified that the sum of the player logits equals to the team logit. This is a numerical illustration of (6) in Theorem 2.1. The "Game Logit" column of Table G5 corresponds to $\text{logit}(p_{gm})$ for player m, $1 \leq m \leq 15$, in game g, $1 \leq g \leq N$, that are used in the $\mathcal{W}(\boldsymbol{X})$ calculation of (12). Prior to performing this calculation, however, all "Active Player" rows (i.e., players with no playing time) are removed. In other words, (12) is calculated over the set $\mathcal{M}_g$, players with playing time in game g, $1 \leq g \leq N$.

TEAM	PLAYER	PTS	FG2O	FG2X	FG3O	FG3X	FTMO	FTMX	PF	AORB	ADRB	STL	BLK	TOV	PFD	SAST	CHGD	AC2P	C3P	DBOX	DFGO	DFGX	DDIS	APM	OCRB	DCRB
PHI	Tobias Harris	18	4	4	3	3	1	1	3	0	1	3	0	0	2	0	0	3	0	1	4	4	1.1	27	1	0
PHI	P.J. Tucker	6	3	0	0	2	0	0	2	1	2	0	1	2	1	2	0	0	2	0	4	5	1.1	21	1	0
PHI	Joel Embiid	26	8	4	1	5	7	2	4	0	9	0	1	6	11	7	0	5	2	4	13	8	1.1	35	1	5
PHI	Tyrese Maxey	21	6	5	2	3	3	0	5	0	1	2	0	1	5	0	0	1	2	0	5	5	1.3	47	0	0
PHI	James Harden	35	4	1	5	4	12	0	3	0	7	0	0	3	4	0	0	1	2	1	11	8	1	66	0	1
PHI	Montrezl Harrell	2	1	2	0	0	0	0	3	0	0	0	1	1	0	0	0	2	1	0	2	3	0.3	9	0	0
PHI	De'Anthony Melton	5	1	1	1	1	0	0	2	0	0	1	0	0	0	0	0	0	1	0	7	1	0.7	23	0	0
PHI	Danuel House Jr.	1	0	1	0	1	1	1	2	0	1	2	0	1	1	0	0	0	2	0	1	3	0.6	9	0	0
PHI	Georges Niang	3	0	1	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0.4	4	0	0
PHI	Matisse Thybulle	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PHI	Team Totals	117	27	19	13	21	24	4	25	1	21	8	3	14	24	9	0	12	12	6	48	38	7.6	241	3	6
BOS	Jaylen Brown	35	10	2	4	8	3	0	1	0	3	2	1	4	3	0	0	2	2	0	5	4	1.2	37	0	0
BOS	Jayson Tatum	35	11	2	2	5	7	2	1	0	12	1	1	3	7	1	0	1	3	0	6	4	1.2	37	0	0
BOS	Al Horford	6	0	2	2	3	0	0	4	0	4	0	0	0	0	1	0	3	1	1	10	8	0.7	26	1	0
BOS	Derrick White	2	1	1	0	1	0	1	2	0	2	1	0	1	3	0	1	5	2	0	4	6	0.9	18	1	0
BOS	Marcus Smart	14	2	2	1	3	7	1	3	0	2	1	0	1	6	0	1	1	1	0	6	4	1.1	44	1	0
BOS	Noah Vonleh	2	1	1	0	0	0	0	4	0	0	1	1	0	0	6	0	1	0	0	3	5	0.6	21	0	2
BOS	Grant Williams	15	2	0	3	0	2	1	3	0	1	0	0	0	2	2	0	1	2	0	6	3	0.7	16	0	0
BOS	Malcolm Brogdon	16	7	2	0	2	2	0	2	0	1	2	0	1	1	0	0	2	3	0	3	8	0.8	22	1	0
BOS	Blake Griffin	1	0	1	0	1	1	1	3	0	2	0	0	0	1	0	0	0	1	1	1	1	0.3	13	2	1
BOS	Sam Hauser	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0.1	0	0	0
BOS	Team Totals	126	34	13	12	23	22	6	24	0	27	8	3	10	23	10	2	16	15	2	45	43	7.6	234	6	3

Table G4: **Illustrative Example: Numerical Counts**. Traditional box score and player tracking data from the October 22, 2022 NBA regular season game between the Philadelphia 76ers (PHI) and Boston Celtics (BOS). Only players with playing time (i.e., `MIN` > 0) are listed.

TEAM	PLAYER	FG2O	FG2X	FG3O	FG3X	FTMO	FTMX	PF	AORB	ADRB	STL	BLK	TOV	PFD	SAST	CHGD	AC2P	C3P	DBOX	DFGO	DFGX	DDIS	APM	OCRB	DCRB	Game Prob.	Game Logit
PHI	Tobias Harris	2.03	2.37	2.18	1.54	-0.23	0.66	1.67	-0.29	-0.69	2.51	-0.31	-0.90	0.70	-0.58	-0.03	1.43	-1.17	0.69	0.85	0.58	0.55	10.27	0.60	-0.51	0.64	0.59
PHI	P.J. Tucker	1.03	-1.63	-0.82	0.54	-1.23	-0.34	0.67	0.71	0.31	-0.49	0.69	1.10	-0.30	1.42	-0.03	-1.57	0.83	-0.31	0.85	1.58	0.55	4.27	0.60	-0.51	0.31	-0.82
PHI	Joel Embiid	6.03	2.37	0.18	3.54	5.77	1.66	2.67	-0.29	7.31	-0.49	0.69	5.10	9.70	6.42	-0.03	3.43	0.83	3.69	9.85	4.58	0.55	18.27	0.60	4.49	0.88	2.01
PHI	Tyrese Maxey	4.03	3.37	1.18	1.54	1.77	-0.34	3.67	-0.29	-0.69	1.51	-0.31	0.10	3.70	-0.58	-0.03	-0.57	0.83	-0.31	1.85	1.58	0.75	30.27	-0.40	-0.51	0.39	-0.45
PHI	James Harden	2.03	-0.63	4.18	2.54	10.77	-0.34	1.67	-0.29	5.31	-0.49	-0.31	2.10	2.70	-0.58	-0.03	-0.57	0.83	0.69	7.85	4.58	0.45	49.27	-0.40	0.49	0.96	3.17
PHI	Montrezl Harrell	-0.97	0.37	-0.82	-1.46	-1.23	-0.34	1.67	-0.29	-1.69	-0.49	0.69	0.10	-1.30	-0.58	-0.03	0.43	-0.17	-0.31	-1.15	-0.42	-0.25	-7.73	-0.40	-0.51	0.15	-1.74
PHI	De'Anthony Melton	-0.97	-0.63	0.18	-0.46	-1.23	-0.34	0.67	-0.29	-1.69	0.51	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-0.17	-0.31	3.85	-2.42	0.15	6.27	-0.40	-0.51	0.13	-1.94
PHI	Danuel House Jr.	-1.97	-0.63	-0.82	-0.46	-0.23	0.66	0.67	-0.29	-0.69	1.51	-0.31	0.10	-0.30	-0.58	-0.03	-1.57	0.83	-0.31	-2.15	-0.42	0.05	-7.73	-0.40	-0.51	0.30	-0.85
PHI	Georges Niang	-1.97	-0.63	0.18	0.54	-1.23	-0.34	-0.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-2.15	-2.42	-0.15	-12.73	-0.40	-0.51	0.20	-1.38
PHI	Matisse Thybulle	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
PHI	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
PHI	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
PHI	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
PHI	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
PHI	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
PHI	Team Total	-2.61	-5.44	0.66	-0.88	5.61	-1.14	5.02	-3.38	-4.31	0.71	-1.65	0.57	4.52	0.37	-0.50	-11.55	-5.57	1.42	0.69	-13.37	-0.69	-9.97	-3.05	-1.69		-2.34
BOS	Jaylen Brown	8.03	0.37	3.18	6.54	1.77	-0.34	-0.33	-0.29	1.31	1.51	0.69	3.10	1.70	-0.58	-0.03	0.43	0.83	-0.31	1.85	0.58	0.65	20.27	-0.40	-0.51	0.69	0.81
BOS	Jayson Tatum	9.03	0.37	1.18	3.54	5.77	1.66	-0.33	-0.29	10.31	0.51	0.69	2.10	5.70	0.42	-0.03	-0.57	1.83	-0.31	2.85	0.58	0.65	20.27	-0.40	-0.51	0.99	4.49
BOS	Al Horford	-1.97	0.37	1.18	1.54	-1.23	-0.34	2.67	-0.29	2.31	-0.49	-0.31	-0.90	-1.30	0.42	-0.03	1.43	-0.17	0.69	6.85	4.58	0.15	9.27	0.60	-0.51	0.14	-1.80
BOS	Derrick White	-0.97	-0.63	-0.82	-0.46	-1.23	0.66	0.67	-0.29	0.31	0.51	-0.31	0.10	1.70	-0.58	0.97	3.43	0.83	-0.31	0.85	2.58	0.35	1.27	0.60	-0.51	0.50	0.02
BOS	Marcus Smart	0.03	0.37	0.18	1.54	5.77	0.66	1.67	-0.29	0.31	0.51	-0.31	0.10	4.70	-0.58	0.97	-0.57	-0.17	-0.31	2.85	0.58	0.55	27.27	0.60	-0.51	0.63	0.52
BOS	Noah Vonleh	-0.97	-0.63	-0.82	-1.46	-1.23	-0.34	2.67	-0.29	-1.69	0.51	0.69	-0.90	-1.30	5.42	-0.03	-0.57	-1.17	-0.31	-0.15	1.58	0.05	4.27	-0.40	1.49	0.55	0.20
BOS	Grant Williams	0.03	-1.63	2.18	-1.46	0.77	0.66	1.67	-0.29	-0.69	-0.49	-0.31	-0.90	0.70	1.42	-0.03	-0.57	0.83	-0.31	2.85	-0.42	0.15	-0.73	-0.40	-0.51	0.66	0.64
BOS	Malcolm Brogdon	5.03	0.37	-0.82	0.54	0.77	-0.34	0.67	-0.29	-0.69	1.51	-0.31	0.10	-0.30	-0.58	-0.03	0.43	1.83	-0.31	-0.15	4.58	0.25	5.27	0.60	-0.51	0.70	0.83
BOS	Blake Griffin	-1.97	-0.63	-0.82	-0.46	-0.23	0.66	1.67	-0.29	0.31	-0.49	-0.31	-0.90	-0.30	-0.58	-0.03	-1.57	-0.17	0.69	-2.15	-2.42	-0.25	-3.73	1.60	0.49	0.49	-0.04
BOS	Sam Hauser	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-0.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-2.15	-3.42	-0.45	-16.73	-0.40	-0.51	0.34	-0.68
BOS	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
BOS	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
BOS	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
BOS	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
BOS	Active Player	-1.97	-1.63	-0.82	-1.46	-1.23	-0.34	-1.33	-0.29	-1.69	-0.49	-0.31	-0.90	-1.30	-0.58	-0.03	-1.57	-1.17	-0.31	-3.15	-3.42	-0.55	-16.73	-0.40	-0.51	0.46	-0.15
BOS	Team Total	4.39	-11.44	-0.34	1.12	3.61	0.86	4.02	-4.38	1.69	0.71	-1.65	-3.43	3.52	1.37	1.50	-7.55	-2.57	-2.58	-2.31	-8.37	-0.69	-16.97	-0.05	-4.69		4.21

Table G5: **Illustrative Example: Centered Statistics and Regression Results**. Centered input data across all players and games for the same October 22, 2022 NBA regular season game between the Philadelphia 76ers (PHI) and Boston Celtics (BOS). This table illustrates Theorem 2.1.

H Performance Measurement Comparisons

The motivation for the flexibility of (10) is a *plug and play* attribute of the proposed ROI framework. For example, it is possible to select any performance measurement of on court basketball performance that is calibrated to a single game for Δ. As we illustrate with Figure 2, this choice can have a significant influence on the dollar allocation of SGV to each player. The purpose of the present section is to provide additional detail on the comparison of player performance for (12), (13), (14), and (15) as it relates to (17).

Figure H1 presents an aggregated comparison of (12), (13), (14), and (15) as it relates to (17) by comparing player percentiles. The off-diagonals show significant disagreements in player performance, especially between PVWL and the other three measures considered. One explanation for these differences is that the model of Table 1 uses player tracking data, which allows for more detail than (2), (3), or BPM (all of which use traditional box score data only). For example, the model of Table 1 does not report assists (AST) as significant but instead finds adjusted passes made (APM) as significant. In comparing PVWS and PVGS, we see general similarities. This may suggest limited differences in these two approaches. PVBPM also reports differences from both PVWS and PVGS. For a summary of the top disagreements between sum totals of (12), (13), (14), and (15), along the lines of (17), see Table H1. For complete results, navigate to the public `github` repository at https://github.com/jackson-lautier/nba_roi.

I Simulation Study

We first conduct a simulation study to verify consistency of the WRMS, (i.e., property (ii) of Theorem 3.1). We assume a sample of $N = 1{,}000$ games, with each team playing between 1 and 5 players (10 total). The number of players appearing for each team is an i.i.d. discrete uniform random variable over the integers $\{1, \ldots, 5\}$. For the performance random variable, we consider two cases. In the first case, we assume the performance random

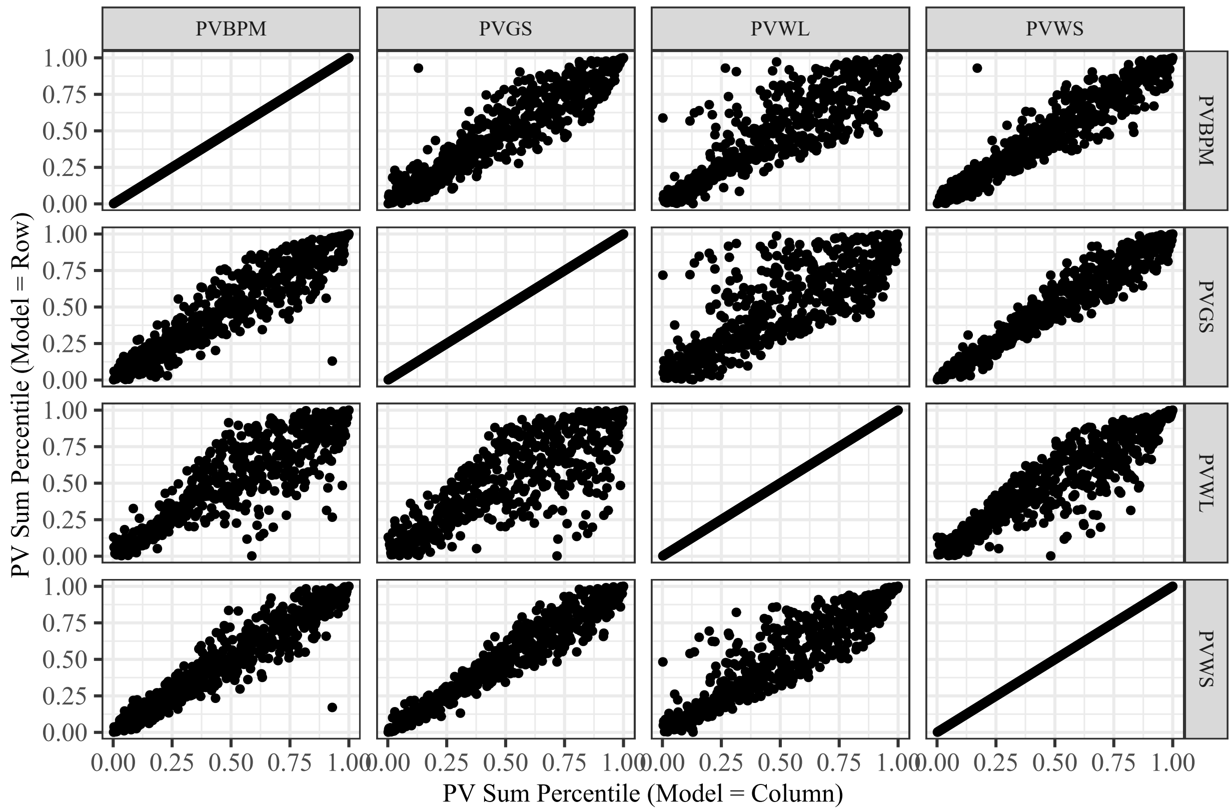

Figure H1: **PVW(·) Percentile Comparisons**. Percentile scatter plots of sum totals of (12) (WL), (13) (GS), (14) (WS), and (15) (BPM) for the 2022-2023 NBA regular season (i.e., (17)) by percentile rank. More deviation from a straight line implies more disagreement between players.

variable for each player follows an i.i.d. exponential distribution with rate parameter equal to 1, denoted exp(1). In the second case, we assume the performance measure random variables are identically distributed normal random variables with a mean of 1 and variance of 2. The

Name	WL(%)	WS(%)	Name	WL(%)	GS(%)	Name	WL(%)	BPM(%)
CJ McCollum	0.31	0.82	Dillon Brooks	0.00	0.72	Wendell Moore Jr.	0.27	0.93
Anfernee Simons	0.16	0.65	Anfernee Simons	0.16	0.85	CJ McCollum	0.31	0.91
Terry Rozier	0.20	0.69	Terry Rozier	0.20	0.87	Dillon Brooks	0.00	0.59
Dillon Brooks	0.00	0.48	Jaden Ivey	0.14	0.80	Jaden Ivey	0.14	0.62
Killian Hayes	0.12	0.54	Jalen Green	0.28	0.92	Trae Young	0.48	0.97
Jaden Ivey	0.14	0.55	CJ McCollum	0.31	0.94	Anfernee Simons	0.16	0.64
Jordan Clarkson	0.21	0.62	Jordan Clarkson	0.21	0.83	Terry Rozier	0.20	0.68
Jalen Green	0.28	0.68	Killian Hayes	0.12	0.72	Jalen Green	0.28	0.73
LaMelo Ball	0.22	0.62	RJ Barrett	0.28	0.84	Killian Hayes	0.12	0.57
Fred VanVleet	0.47	0.86	LaMelo Ball	0.22	0.76	Fred VanVleet	0.47	0.91
Name	WS(%)	BPM(%)	Name	GS(%)	BPM(%)	Name	WS(%)	GS(%)
Wendell Moore Jr.	0.17	0.93	Wendell Moore Jr.	0.13	0.93	Jordan Poole	0.66	0.91
Steven Adams	0.83	0.49	Sam Hauser	0.56	0.90	Jaden Ivey	0.55	0.80
Jusuf Nurkic	0.83	0.53	Dyson Daniels	0.42	0.75	Jalen Green	0.68	0.92
AJ Griffin	0.55	0.81	Zion Williamson	0.75	0.45	Dillon Brooks	0.48	0.72
Isaiah Stewart	0.73	0.47	Luke Kornet	0.49	0.78	Isaiah Hartenstein	0.87	0.65
Anthony Gill	0.38	0.63	LaMelo Ball	0.76	0.47	Andre Drummond	0.79	0.57
Jalen Duren	0.92	0.67	Anthony Gill	0.35	0.63	Jordan Clarkson	0.62	0.83
Precious Achiuwa	0.68	0.44	John Konchar	0.55	0.84	Steven Adams	0.83	0.63
Sam Hauser	0.66	0.90	Cam Thomas	0.55	0.28	Usman Garuba	0.65	0.45
Devonte' Graham	0.54	0.79	Matisse Thybulle	0.47	0.75	Anfernee Simons	0.65	0.85

Table H1: **Player Performance Disagreements**. The top ten largest disagreements between sum totals of (12) (WL), (13) (GS), (14) (WS), and (15) (BPM) for the 2022-2023 NBA regular season (i.e., (17)) in terms of percentile rank (%).

covariance matrix is as follows:

$$\Sigma = \begin{bmatrix} 2.00 & 0.36 & 0.50 & 0.60 & 0.71 & -0.73 & -0.70 & -0.48 & -0.32 & -0.73 \\ 0.36 & 2.00 & 0.17 & 0.22 & 0.22 & -0.64 & -0.61 & -0.39 & -0.21 & -0.79 \\ 0.50 & 0.17 & 2.00 & 0.76 & 0.71 & -0.76 & -0.58 & -0.41 & -0.38 & -0.16 \\ 0.60 & 0.22 & 0.76 & 2.00 & 0.50 & -0.25 & -0.42 & -0.48 & -0.42 & -0.74 \\ 0.71 & 0.22 & 0.71 & 0.50 & 2.00 & -0.59 & -0.40 & -0.31 & -0.19 & -0.19 \\ -0.73 & -0.64 & -0.76 & -0.25 & -0.59 & 2.00 & 0.70 & 0.52 & 0.29 & 0.26 \\ -0.70 & -0.61 & -0.58 & -0.42 & -0.40 & 0.70 & 2.00 & 0.23 & 0.38 & 0.73 \\ -0.48 & -0.39 & -0.41 & -0.48 & -0.31 & 0.52 & 0.23 & 2.00 & 0.88 & 0.39 \\ -0.32 & -0.21 & -0.38 & -0.42 & -0.19 & 0.29 & 0.38 & 0.88 & 2.00 & 0.36 \\ -0.73 & -0.79 & -0.16 & -0.74 & -0.19 & 0.26 & 0.73 & 0.39 & 0.36 & 2.00 \end{bmatrix}.$$

The simulation procedure is

1. Simulate either (case 1) 1,000 $\times$ 10 i.i.d. exp(1) random variables or (case 2) 1,000

dimension 10 multivariate normal (MVN) random variables, MVN($\mathbf{1}_{10}, \Sigma$).

2. For each game, $g = 1, \ldots, 1{,}000$, simulate two discrete uniform random variables over $\{1, \ldots, 5\}$ to determine how many players appear for each team.

3. For each game, $g = 1, \ldots, 1{,}000$, calculate the natural share, as defined by (7), using either (case 1) the simulated i.i.d. exp(1) random variables or (case 2) the MVN($\mathbf{1}_{10}, \Sigma$) random variables from Step 1.

4. For each player, $m \in \mathcal{M}_g$, appearing in each game, g, $1 \leq g \leq 1{,}000$, we calculate $\mathcal{W}$.

5. For each player, $m \in \mathcal{M}_g$, appearing in each game, g, $1 \leq g \leq 1{,}000$, we calculate the bias by subtracting the calculated natural share in Step 3 from the calculated $\mathcal{W}$ in Step 4.

From our sample, we obtain $m^* = 6{,}081$, $\bar{m} = 6.081$, and (case 1) $\bar{\Delta}_{m^*} = 0.9939$, $s(\Delta)_{m^*} = 0.9861$ or (case 2) $\bar{\Delta}_{m^*} = 1.0050$, $s(\Delta)_{m^*} = 1.4195$. This results in an empirical mean bias of 0.0000 (both case 1 and case 2) over the simulated sample of 6,081 players (the empirical median bias is 0.0007 for case 1 and 0.042 for case 2). This is numerical verification of Theorem 3.1, (ii).

We next provide a simulation study to verify the results of Theorem 3.2. We estimate (16) using (12) for all $g = 1, \ldots, n/2$ and $\pi \in \mathcal{P}$ using data from the 2022-2023 NBA regular season. These estimates correspond to Section 3.1. Thus, $n = 2{,}452$. Further, we assume $\mathrm{SGV}_g \sim \mathcal{N}(\mu = 100, \sigma^2 = 25)$ 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 $\widehat{\mathrm{SGV}}_g$, $g = 1, \ldots, 1{,}226$.

2. Compute the product

$$\hat{\theta}_g = \widehat{\mathrm{SGV}}_g \sum_{\pi \in \overline{\mathcal{M}}_g} \mathcal{W}(\boldsymbol{X})^*_{g\pi},$$

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

3. Save the result as the summation,

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

In doing so, we find an empirical mean of

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

which is quite close to $\mu(n/2) \equiv 100 \times 1{,}226$. In Figure I1, we provide a density plot of the simulated results.

Next, we state a minor extension to Theorem 3.2.

Result C.1. *Assume the conditions of Theorem 3.2, and further assume* $\text{Var}(\text{SGV}_g) = \sigma^2$ *for all* $g = 1, \ldots, N \equiv n/2$. *If* SGV_g *is independent of* SGV_{g^*} *for all* $g, g^* = 1, \ldots, n/2$, $g \neq g^*$, *then*

$$\text{Var}\left(\sum_{g=1}^{n/2} \sum_{\pi \in \overline{\mathcal{M}}_g} \text{SGV}_g \mathcal{W}^*_{g\pi} \middle| \mathcal{W}^*_{g\pi} \right) = \sigma^2 \sum_{g=1}^{n/2} \left(\sum_{\pi \in \overline{\mathcal{M}}_g} \mathcal{W}^*_{g\pi} \right)^2.$$

Proof. By independence,

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

□

In an additional simulation study with 10,000 replicates, we obtain an empirical sample

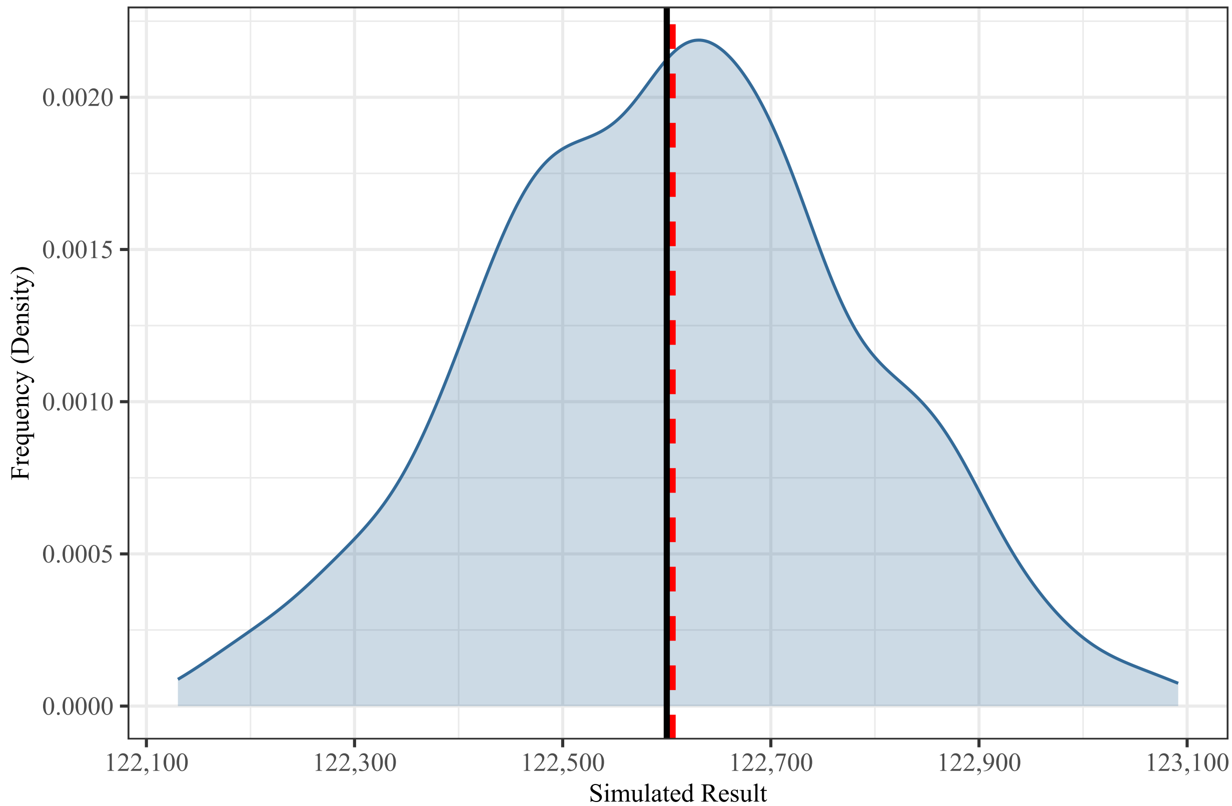

Figure I1: **Simulation Study Results**. A density plot of 1,000 replicates to verify Theorem 3.2. The vertical black line indicates the theoretical mean using Theorem 3.2. The vertical dashed line indicates the empirical sample mean of the 1,000 replicates. The two quantities are quite close, which is a simulation validation of Theorem 3.2.

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.

Finally, we verify the results of Section E with a simulation study. In this instance, we assume a sample of $N = 1{,}000$ games, with each team playing a nonrandom 5 players. The number of players is held fixed to verify the results of Section E. Further, we assume the i.i.d. performance random variables are $\Delta = -0.25\rho_1 + 0.25\rho_2$, where $\rho_1 \sim \mathcal{N}(\mu = 0, \sigma = 5)$ and $\rho_2 \sim \mathcal{N}(\mu = 0, \sigma = 7)$. Thus, the natural share defined in (7) follows (S.1) with $\sigma_x^2 = 5^2/16 + 7^2/16 = 4.625$ and $\sigma_y^2 = 9\sigma_x^2$. To verify this with simulation, we

1. Simulate $1{,}000 \times 10$ i.i.d. $\Delta = -0.25\rho_1 + 0.25\rho_2$ random variables.

2. For each game, $g = 1, \ldots, 1{,}000$, calculate the natural share, as defined by (7), using

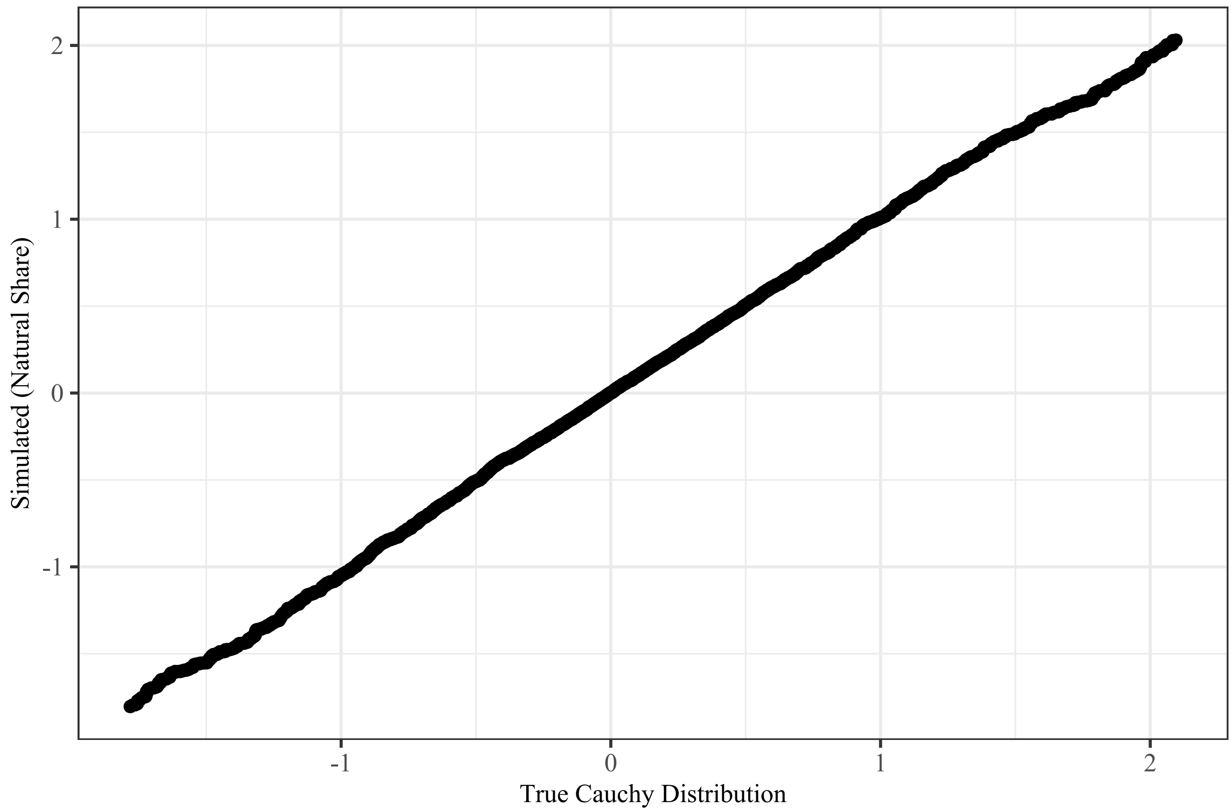

Figure I2: **Cauchy Simulation Results**. A QQ-plot of the middle 90% of ordered data from simulated natural shares in the form of a ratio of independent normal random variables and a Cauchy distribution with location and scale parameters per (S.1). The closeness of the distributions represents simulation verification of the result of Section E.

the simulated i.i.d. exp(1) random variables from Step 1.

3. Simulate 10,000 Cauchy random variables with location parameter $x_0 = 0.10$ and scale parameter $\gamma = 0.3$ per (S.1).

4. Compare a QQ-plot of the middle 90% of the ordered 10,000 observations from Step 2 and the ordered 10,000 observations from Step 3. We use the middle 90% because of the tendency for extreme observations from the Cauchy distribution. The results may be found in Figure I2, which indicates numerical validation of the result of Section E.

Team	# Households
ATL	2,679,850
BKN	7,726,580
BOS	2,596,190
CHA	1,323,400
CHI	3,624,820
CLE	1,552,420
DAL	3,041,540
DEN	1,792,540
DET	1,937,250
GSW	2,593,210
HOU	2,666,330
IND	1,207,280
LAC	5,838,090
LAL	5,838,090
MEM	644,360
MIA	1,720,970
MIL	900,200
MIN	1,839,480
NOP	687,110
NYK	7,726,580
OKC	743,340
ORL	1,775,140
PHI	3,108,960
PHX	2,138,870
POR	1,293,400
SAC	1,502,080
SAS	1,059,540
TOR	8,297,000
UTA	1,148,120
WAS	2,617,350

Table J1: **Local Television Market Size by # Households**. Estimates from Nielsen (2022). Local television revenue assumed to be $1 per household.

J Local Television Market Estimates

Table J1 provides a complete list of the number of television households by market for each team (Nielsen, 2022). In the Single Game Value estimates, the $\boldsymbol{\phi}$ parameter estimates are assumed to be $1 per household. For example, the estimate of $\hat{\phi}_{\text{ATL}}$ = $2,679,850.

References

Basketball for Coaches (2024). “How to box out in basketball (4-step guide).” https://www.basketballforcoaches.com/ - Basketball for Coaches. https://www.basketballforcoaches.com/box-out-basketball/. Online; accessed 20 January 2025.

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*. Stanford 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.

E. Lehmann and G. Casella (1998). *Theory of Point Estimation, 2nd Edition*. Springer.

J. MacMullan (2015). “How Tim Duncan and the Spurs’ majestic 2014 Finals changed the NBA forever.” https://www.ESPN.com - Entertainment and Sports Programming Network. `https://www.espn.com/nba/playoffs/2015/story/_/page/PresentsSpursHeat/how-spurs-2014-finals-performance-changed-nba-forever`. Online; accessed 17 January 2025.

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

J. Martinéz (2019). "A more robust estimation and extension of factors determining production (FDP) of basketball players." **6**, 81–85.

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.

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 (2023a). "Glossary." nba.com. https://www.nba.com/stats/help/glossary. Online; accessed 13 June 2023.

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

Nielsen (2022). "2022-2023 local television market universe estimates." National Media Spots. https://www.nationalmediaspots.com/media-stats/Nielsen-2022-2023-DMA-Ranks.pdf. Online; accessed 11 February 2025.

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

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.