Research Article

Andrew J. Cannon*, Jared D. Fisher, Gilbert W. Fellingham and Garritt L. Page

Analyzing the effects of NBA head coaches

Abstract: Evaluating the value-added of coaches in the NBA is a challenging task as the coaches with the best win/loss records often have the best players. This prompts a question of attribution: if two coaches had the same roster, which one would win? This paper attempts to answer this question by introducing a method for quantifying coaching effect in the NBA. We propose a method for isolating the effect of a coach's in-game scheme on their team's probability of winning a game while controlling for other factors, namely the relative strength of the two competing teams. To control for team strength, player performance metrics are aggregated into "Team-Adjusted VORP Difference" or ΔtVORP, meant to account for the difference in quality of on-court product between both teams. We model each coach's win probability as a function of ΔtVORP using probit monotone Bayesian Additive Regression Trees. In comparing coaches' win probability curves, we find some of the winningest coaches are close to average in terms of scheme, while other coaches are found to be truly great contributors to their teams.

Keywords: Bayesian additive regression trees; coaching; machine learning; monotonic function estimation; basketball

1 Introduction

Coaches in the National Basketball Association (NBA) are paid millions of dollars every year to guide their teams to successful seasons. Gregg Popovich was one of the longest tenured head coaches in the NBA and made up to $16 million per year, while a coach who makes the median coach salary is paid between $4 and $6 million depending on the year. This salary is similar to the median player salary (just over $4 million). Given the financial emphasis placed on having a high quality coach, it is reasonable to ask if they are worth the large investment. This question requires more than simply looking at a head coach's win percentage or number of championships as there are many different factors which contribute to team success and more specifically to winning games.

Many attempts have been made to evaluate the value of coaches to their sports teams. Much of the existing literature focuses on determining whether coaches truly are to blame for the success or failure of their teams. Dawson et al. (2000) applies stochastic frontier analysis to quantify coaching effect in professional English association football. They maximize the attainable level of team performance by coaches given the available playing talent level using a linear model. In short, they use a logistic regression of win percentage on a team performance metric. Berry and Fowler (2019) propose a method to estimate the variation in team and player metrics contributed directly by a coach in various professional sports. They determine that coaches explain up to 30 % of the variation in points scored and allowed in both professional and college basketball. They note this effect to be quite large when compared with other professional sports.

Specifically in the NBA, Hofler and Payne (2006) estimate teams' winning potential using a stochastic production frontier model and compare actual wins to this estimate. They attribute the positive or negative difference in actual and estimated wins partially to coaching quality. Fort et al. (2008) evaluate coaching effectiveness using a measure formed from player and team offensive and defensive factors they call technical efficiency. It is used to give coaches a season level "score" on their effectiveness and then these scores are compared across coaches. While other works have evaluated NBA coaches in a different ways, this paper tries to estimate what a coach contributes to the number of wins their team gets in a season beyond what their roster contributes. It estimates a single number for each coach for how many wins or losses above average each coach is worth.

1.1 Components of coaching

For a specific game g, we propose that an NBA coach's effect on a team can be divided into three different categories.

*Corresponding author: Andrew J. Cannon, Department of Statistics, Iowa State University, Ames, IA, USA, E-mail: werd6432@iastate.edu
Jared D. Fisher, Gilbert W. Fellingham and Garritt L. Page, Department of Statistics, Brigham Young University, Provo, UT, USA. https://orcid.org/0000-0002-0369-9203 (J. D. Fisher)

1. **Long-term effects**, such as player development, injury management, team chemistry, personality management, "culture", and anything else that cannot be changed during the week or day of game g.
2. **Playing time**, specifically the time allotted to each player in game g. Naturally, players who are not available to play due to injury or other matters have zero playing time.
3. **Scheme**, the coach's strategy for the players he has chosen to put on the court. Colloquially called "X's and O's", this includes the team's defensive system, play calling, in-game manipulation, and direction that a coach gives to his/her players.

This partitioning of a coach's effect makes the argument that there are some things a coach can and cannot change for a single game, and we assume that "playing time" and "scheme" are the two things a coach can change for a single game.

We present a method for evaluating a coach's scheme when controlling for the other factors that affect a single game. We illustrate this in Figure 1. We see that, in a single game, a coach affects the scheme and the playing time allotted to players. The coach cannot change the skill and ability of the team's players on the day of the game; player development is a long term endeavor. The combination of playing time and the strength/ability of the players on the roster is what we call "on-court product". We separate on-court product from coach's scheme as even with no coach the players

could assign playing time to usual rotations, whereas specific schemes are implemented by coaches. We recognize that this does not account for the opposing coach's scheme. Furthermore, as we are only interested in a coach's individual effect, we do not include interaction effects between coaches.

Following the graph in Figure 1, in order to evaluate the effect of a coach's scheme we need to particularly consider two functions. First, we must consider how to combine playing time allotment with roster strength, i.e. individual player skills, to measure a team's on-court performance. To do this, we propose a new metric that we call "Team-Adjusted VORP" which accounts for player contributions and minutes played. Second, we must consider how to model win probability based on a coach's scheme and the on-court products of both teams. Each coach can be modeled separately by regressing win/loss on the difference in their tVORP and opponent tVORP. We propose that this regression can be done with monotone Bayesian additive regression trees to provide the right balance between structure and flexibility.

The rest of this paper proceeds as follows. Section 2 describes the creation of tVORP. Section 3 describes the probit monotone BART model (Fisher 2025) that we use to model expected win percentage conditional on coach and ΔtVORP, which is the difference between teams' tVORP. Section 4 presents the results. Section 5 presents discussion and concludes.

2 Measuring player contributions

Quantifying player value and skill has been an important focus of sports analytics in the NBA. Along with basic box score metrics, more advanced metrics have been developed to standardize across position and game situation and compare to an "average" player. Value Over Replacement Player (VORP), which is detailed in Section 2.2, is commonly used across literature to assess player strength and predict outcomes. Wu et al. (2018) used VORP to predict NBA player salaries. Chen et al. (2019) and McCorey (2021) found VORP, among other box score metrics and more advanced metrics, to be significant in predicting the MVP in the NBA. VORP gives a good snapshot of box score contribution and also is significant in predicting things outside game statistics like player salary, making it a good measure of overall player skill and value to their team.

In Section 2.3 we describe our new proposed metric "tVORP" that combines a player's season-level VORP with their minutes allotted in a specific game to get a measure of the team-level on-court product. tVORP is constructed using
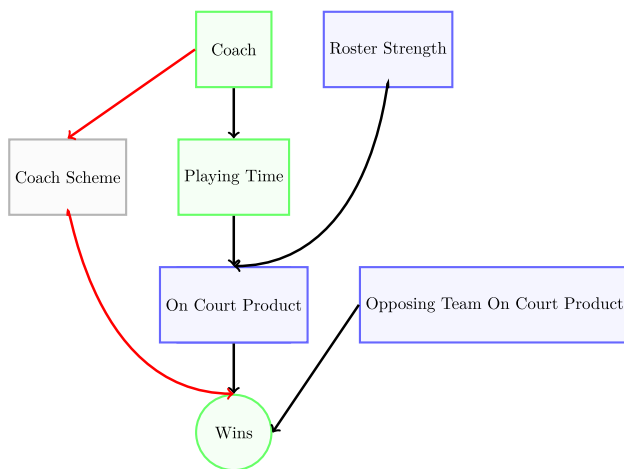

Figure 1: A graph illustrating the variables we are considering here and their effects on winning. Gray rectangles indicate variables which are unobserved. Green rectangles represent variables which can be measured. Blue rectangles represent variables which cannot be directly quantified, but we do have reasonable metrics which can approximate them. The gray rectangle "Coach Scheme" represents the function that will be quantified using $f_{coach}(\Delta tVORP)$.

season-level strength for each player. It is meant to account for the contribution of a player to their team relative to their team strength over the entire season. Coach scheme is implemented on a game-to-game basis as different opponents require different game plans. Since tVORP is calculated using season-level VORP instead of game-level VORP, it is "averaged" over all the different game plans used for each game.

2.1 BPM: Box Plus/Minus

From Figure 1, on-court product is a function of player season-level strength and minutes allotted. Winning is a function of on-court product, coach scheme, and opposing team on-court product. Measuring on-court product and opposing on-court product allows us to approximate the effect of coach scheme through modeling. The players who are available on any given roster and the minutes allocated to each are publicly available data that can be accessed through box scores. All data used in this project are from basketball-reference.com, which has publicly available data for every NBA game and season since 1946. Figure 2 shows an example of a basic box score.

Box score statistics can account for different portions of player performance. For example, field goals made (FG) is one possible measure for the offensive output of a player during the game, but doesn't include volume of scoring attempts (field goals attempted), quality of passes to teammates (assists), etc. To more effectively capture all aspects of a player's game, Box Plus/Minus (BPM) was introduced by Daniel Myers before the 2007–2008 season, then retroactively calculated for past seasons (Myers 2020). BPM is described as an estimate of a basketball player's full on-court contribution to the team and is calculated using the process described below.

"Raw BPM" is calculated as a linear combination of box score statistics:

$$\text{Raw BPM}_i = x_i^T \beta_i + \text{Adjustment Constant}$$

The box score statistics x_i are multiplied by coefficients β_i specific to player's estimated position. Position is empirically estimated in a spectrum from pure point guard to pure center, such that it is in [1, 5]. Myers (2020) uses 2017 LeBron James as an example to help explain the process, which we reproduce in Table 1, where the coefficients shown represent LeBron's estimated position of 2.3 (as opposed to a "pure small forward" at 3.0) such that the linear combination yields a value of 18.7. A constant is then added to this value (like an intercept) to adjust for both estimated position and estimated offensive role (as a spectrum from "creator" to "receiver"). For LeBron James in 2017, this adjustment constant is −3.1, resulting in a Raw BPM of 18.7 − 3.1 = 15.6. For further detail on the coefficient values and adjustment constants, see Myers (2020).

	Basic Box Score Stats																	
Starters	**MP**	**FG**	**FGA**	**FG%**	**3P**	**3PA**	**3P%**	**FT**	**FTA**	**FT%**	**ORB**	**DRB**	**TRB**	**AST**	**STL**	**BLK**	**TOV**	**PF**	**PTS**	**+/-**
Rajon Rondo	43:33	9	14	.643	0	2	.000	2	4	.500	0	7	7	13	0	0	4	4	20	-13
Paul Pierce	40:34	6	15	.400	2	4	.500	9	9	1.000	0	5	5	5	2	0	0	3	23	-17
Kevin Garnett	31:43	4	8	.500	0	0		1	1	1.000	1	11	12	2	0	2	5	4	9	-4
Brandon Bass	28:19	6	11	.545	0	0		3	4	.750	6	5	11	1	0	0	1	2	15	-8
Courtney Lee	24:00	5	6	.833	1	1	1.000	0	0		0	1	1	1	0	0	1	5	11	-7
Reserves	**MP**	**FG**	**FGA**	**FG%**	**3P**	**3PA**	**3P%**	**FT**	**FTA**	**FT%**	**ORB**	**DRB**	**TRB**	**AST**	**STL**	**BLK**	**TOV**	**PF**	**PTS**	**+/-**
Jason Terry	24:51	2	7	.286	0	3	.000	4	4	1.000	0	0	0	1	1	0	3	3	8	-10
Jeff Green	23:04	0	4	.000	0	0		3	4	.750	0	3	3	0	1	0	1	0	3	-7
Leandro Barbosa	15:50	6	8	.750	3	3	1.000	1	2	.500	0	1	1	1	0	0	0	1	16	+4
Jared Sullinger	8:06	1	2	.500	0	0		0	0		0	1	1	0	0	0	0	1	2	-3
Jason Collins							Did Not Play													
Kris Joseph							Did Not Play													
Darko Miličić							Did Not Play													
Chris Wilcox							Did Not Play													
Team Totals	**240**	**39**	**75**	**.520**	**6**	**13**	**.462**	**23**	**28**	**.821**	**7**	**34**	**41**	**24**	**4**	**2**	**15**	**23**	**107**	

Figure 2: Box score from the Miami Heat's regular season defeat of the Boston Celtics on October 30, 2012. It includes game metrics for all Celtics players who participated in that game. This graphic was taken directly from basketball-reference.com.

Table 1: Per 100 possessions box score metrics for LeBron James in 2017 with the corresponding regression coefficients used to calculate raw season-level BPM (Myers 2020).

Variable	Coefficient	Per 100 possession statistics (2017 LeBron)	Total
Points (adjusted for team context)	0.860	34.9 adjusted to 30.4	26.1
3-Pointers made	0.389	2.2	0.9
Assists	0.727	11.5	8.4
Turnovers	−0.964	5.4	−5.2
Offensive Rebounds	0.473	1.7	0.8
Defensive Rebounds	0.137	9.7	1.3
Steals	1.252	1.6	2.0
Blocks	1.125	0.8	0.9
Personal Fouls	−0.367	2.4	−0.9
Field goals attempted	−0.560	24.0	−13.4
Free Throws attempted	−0.246	9.5	−2.3
Total			18.7

Raw BPM is next adjusted to account for team strength. Team adjusted rating, notated α_j, for team j is calculated using offensive efficiency (in points per 100 possessions) and average points ahead. The negative effect of leading on offensive performance was first examined by Goldman and Rao (2013) and further quantified by Engelmann (2014). It was found that, on average, a team plays 0.35 pts worse per 100 possessions for every point of lead. Thus, the offensive efficiency is increased or decreased accordingly.

$$\alpha_j = \text{OffensiveEfficiency}_j + \frac{-0.35 * \text{AverageLead}_j}{2} \quad (1)$$

The effect of team lead is allocated half to the team in the lead and half to the opposing team, hence the division by two in Equation (1). For example, if team 1 had an offensive efficiency of +3.0 and an average lead of 1.4, α_1 would be 2.755.

The sum of all players' Raw BPM weighted by possessions played, multiplied by the team adjustment constant notated ω_j, must equal α_j. This adjusts a player's BPM through the "residual" of the combination of playing time and Raw BPM on α_j as shown in Equation (2). For players $i = 1, 2,\ldots, n_j$ on team j:

$$\alpha_j = \left(n_j \cdot \omega_j\right) + \sum_{i=1}^{n_j} \left[\%\text{PossessionsPlayed}_i \cdot \text{RawBPM}_i\right]$$

$$\Rightarrow \omega_j = \frac{\alpha_j - \sum_{i=1}^{n_j} \left[\%\text{PossessionsPlayed}_i \cdot \text{RawBPM}_i\right]}{n_j}$$

$$(2)$$

Thus the team adjustment constant is calculated using α_j and the weighted sum of Raw BPM. ω_j shifts the sum of each player's weighted Raw BPM. This value can be seen as an intercept in the linear regression model for each player as shown in Equation (3).

$$\text{BPM}_i = \omega_j + \text{RawBPM}_i \quad (3)$$

For example, if we hold α_j constant, higher Raw BPM across the team means ω_j will be more negative. Holding Raw BPM across the team constant, if α_j increases then ω_j will be less negative. The purpose of ω_j then is to scale individual player box score performance back to what the team is able to accomplish together in terms of efficiency and lead.

Altogether, BPM is able to account for player strength in terms of what they produce on their team in the box score. However, one of the main weaknesses of BPM as a player evaluation metric: it does not account for how much a player actively plays in a season. We will account for this weakness by using VORP.

2.2 VORP

As described above, one weakness of BPM is that it is purely a rate statistic. It does not take into account playing time and number of games played in a season. Woolner (2002) introduced the idea of value over replacement player (VORP) applied to baseball. VORP, when applied to basketball, converts BPM into an estimate of each player's overall contribution to their team over a replacement level player. As described by Myers (2020), for player i on team j in season h, it is calculated:

$$\text{VORP}_{ih} = [\text{BPM}_{ih} - (-2.0)] \cdot \left[\%\text{PossessionsPlayed}_{ih} \cdot \frac{\text{TeamGames}_{jh}}{82}\right] \quad (4)$$

The accepted "replacement player" is defined as a player on minimum salary for a team. BPM level for this player across the NBA is computed to be -2.0 by (Myers 2020), thus we subtract negative two from BPM to equate to a replacement player. This quantity is then multiplied by the percentage of possessions played in the season in order to take a player's season-level usage into account. This is further multiplied by the ratio of games played to adjust for shortened seasons (82 games is standard). VORP has been calculated in the same way since 1985, so for the sake of consistency the data used in this study are limited to regular season games from 1985 to 2019.

2.3 tVORP

We propose a team-level measure of roster strength aggregated and weighted at the game-specific level that we call Team-Adjusted VORP (tVORP). Its construction will be described here. Roster strength is considered in this paper as the weighted sum of all the available individual player strengths. As shown in Figure 1, both roster strength and playing time factor into on-court product. Thus, for players $i = 1, 2, \dots n$ playing in game g:

$$tVORP_{jhg} = \frac{\sum_{i=1}^{n} VORP_{ih} \cdot GameMinutes_{ig}}{240} \qquad (5)$$

As shown above in Equation (5), VORP is multiplied by the number of minutes allocated to that player by the coach in a game, then is divided by the total number of minutes available to be allocated to players in a game. The divisor is calculated by multiplying the 48 min in every game by the 5 players on the floor at any given time. The estimated density of tVORP is shown in Figure 3.
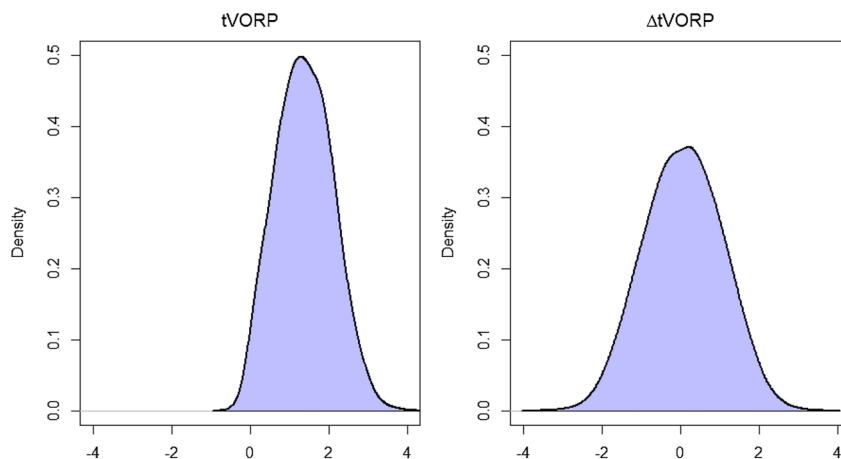
2.4 ΔtVORP

tVORP can vary widely from game to game. For example, a team could win a game with a tVORP as high as 4.0 or as low as -4.0, as the competition also depends on the strength of their opponent. Thus, we will use tVORP difference (ΔtVORP), being the difference in tVORP between the team of interest and their opponent, as the independent variable in this analysis. The density of ΔtVORP, as shown in Figure 3, is centered around zero and relatively symmetric. This allows us to interpret significant positive deviations from zero in ΔtVORP as the team of interest's on-court product being higher quality than their opponents. This applies conversely in the negative deviation direction as well.

In summary, VORP is a measure of player season-level strength. tVORP takes playing time into account, thus is a measure of roster strength on a particular day. Lastly, the strength of the opposing team is accounted for when ΔtVORP is calculated. Therefore, ΔtVORP is an approximation for the on-court product and opposing team on-court product shown in Figure 1.

3 Modeling

As shown in Figure 1, we need to appropriately control for on-court product in order to compare coaches' scheme. ΔtVORP was introduced as our approximation for the difference between teams' on-court product. We now propose coach-specific functions $f_{coach}(\Delta tVORP)$ which allow us to model winning as follows:

$$Win \sim Bernoulli(f_{coach}(\Delta tVORP)). \qquad (6)$$

Figure 3: Density of tVORP and ΔtVORP. We see that ΔtVORP is normally distributed around zero. This illustrates that large deviations from ΔtVORP = 0, or games where a coach's on-court product is much better or worse than the opposing coach's on-court product, are much less common than games where the team strengths are similar.

We can then compare coaches' schemes by comparing their expected win probabilities $f(\Delta tVORP)$ for various values of $\Delta tVORP$.

A natural starting point for modeling binary outcomes like wins/losses is logistic regression. However, we have found that these types of curves need more flexibility than the two parameters in standard logistic regression. Thus we propose using Bayesian additive regression trees, or BART (Chipman et al. 2010). BART for binary outcomes is typically constructed using a probit link. However, we believe the relationship between the difference in team strength ($\Delta tVORP$) and the expected winning percentage for a coach should be nondecreasing, akin to how a standard logistic regression would fit. Chipman et al. (2022) introduces constrained priors that create monotone BART for continuous outcomes, and Fisher (2025) expands monotone BART to work with binary outcomes by adding a probit link.

We propose using probit monotone BART of Fisher (2025) to estimate f_{coach}. To introduce these ideas, we first detail probit BART of Chipman et al. (2010) in Section 3.1, then in Section 3.2 we highlight how Fisher (2025) enhances probit BART with the monotonic constraints of Chipman et al. (2022).

3.1 Probit BART

Following Chipman et al. (2010), probit BART regresses binary outcome Y_i on vector of covariates $x_i \in \mathbb{R}^P$ using tree models with a probit link. In other words, for $i = 1, \ldots, n$,

$$P[Y_i = 1|x_i] = \Phi[G(x_i) + c] \tag{7}$$

$$G(x) = \sum_{j=1}^{m} g(x; T_j, M_j) \tag{8}$$

where G is a sum of m regression trees g. If interested in shrinking $P[Y_i = 1|x_i]$ to a value other than 0.5, we can set c to a value other than 0, typically $c = \Phi^{-1}(\bar{y})$.

Chipman et al. (2010) parameterize each regression tree $g(x; T_j, M_j)$ with T_j and M_j, following Chipman et al. (1998). T_j defines the tree structure, i.e. where there are binary splits in the domain of x. The terminal leaf nodes $\mu_{\ell j}$, $\ell \in \{1, \ldots, b_j\}$ of each tree return a value for the function $g()$, and these leaf node values make up the vector M_j. Each tree's priors are assumed independent, as are and the priors for the $\mu_{\ell j}$ conditional on the tree structure T_j, which yields

$$p((T_1, M_1), \ldots, (T_m, M_m)) = \prod_{j=1}^{m} \left[p(T_j) p(M_j|T_j) \right] \tag{9}$$

$$p(M_j|T_j) = \prod_{\ell}^{b_j} p(\mu_{\ell j}|T_j). \tag{10}$$

The prior for each tree structure/partition T_j is defined by three probabilities: the probability that a node is nonterminal, the probability of which covariate is chosen for a split, and the probability/distribution of what value of that covariate to split on. The splitting variable and location are given uniform priors, while the probability a node is nonterminal at tree depth d is specified as $\alpha(1 + d)^\beta$. The priors for the leaf nodes $\mu_{\ell j}$ given the tree structure T_j are chosen to be normal with a particular variance

$$\mu_{\ell j} \sim N\left(0, \sigma_\mu^2\right) \tag{11}$$

$$\sigma_\mu = 3/(k\sqrt{m}) \tag{12}$$

where k controls the degree of shrinkage and is typically chosen to put large prior probability that $(G(x) + c) \in (-3, 3)$.

Chipman et al. (2010) find that the default hyperparameter values they propose work well in a variety of circumstances. They recommend $\alpha = 0.95$, $\beta = 2$, $k = 2$ and typically $m = 50$ or $m = 200$.

3.2 Probit monotone BART

The probit monotone BART model from Fisher (2025) evolves from the original probit BART framework by incorporating the monotonicity constraints as introduced in Chipman et al. (2022). As a sum of nondecreasing functions will itself be nondecreasing, we can simply constrain the individual trees to be monotonic functions. This means two main changes to the original probit BART. First, instead of the prior in Equation (10), we set

$$p(M_j|T_j) \propto \left[\prod_{\ell}^{b_j} p(\mu_{\ell j}|T_j) \right] \chi_C(T_j, M_j) \tag{13}$$

where χ is an indicator function equal to 1 when $(T_j, M_j) \in C$, and 0 otherwise. Chipman et al. (2022) define C as the set of all (T_j, M_j) which satisfy the monotonicity constraints:

$$C = \{(T, M): g(x; T, M) \text{ is monotone for each } x_p \in S\} \tag{14}$$

where S is the subset of all covariates $\{1, \ldots, P\}$ that are constrained to have monotonic relationships with $P[Y_i = 1|x]$. This means that, in MCMC, the terminal node values $\mu_{\ell j}$ are sampled from distributions truncated by their neighboring nodes such that $g(x; T, M)$ is nondecreasing in the desired covariates.

Second, Fisher (2025) follows Chipman et al. (2022) and tweaks the prior on μ in Equation (11) into

$$(\mu_{\ell j}|T_j) \sim N\left(0, c^2\sigma_\mu^2\right) \tag{15}$$

by scaling the typical variance by a constant $c^2 = \frac{\pi}{\pi-1} \approx 1.467$. This c^2 is specifically chosen so that the marginal variance of each $\mu_{\ell j}$ in small trees is again σ_μ^2 (marginal in the sense of no monotonic constraints from the neighboring leaf nodes).

For the hyperparameter values, Fisher (2025) chooses $m = 200$ trees, $k = 2$, $\alpha = 0.25$ and $\beta = 0.8$. While 200 trees and $k = 2$ are common practice with BART, these particular choices of α and β are used by Chipman et al. (2022) to yield

monotonic-constrained trees that are about as large as trees from standard BART.

4 Results

Using probit monotone BART (mBART), we model $f_{coach}(\Delta tVORP)$ for each coach using $\Delta tVORP$ as the sole input variable and each game's win/loss as our outcome

(a) Del Harris Led his teams to the playoffs eleven of fourteen seasons. Hall of Fame inductee.

(b) Gregg Popovich was the longest tenured head coach in history. Led his team to five NBA championships. Hall of Fame inductee.

(c) Phil Jackson led his teams to eleven NBA championships, the most in NBA history for a coach. Hall of Fame inductee.
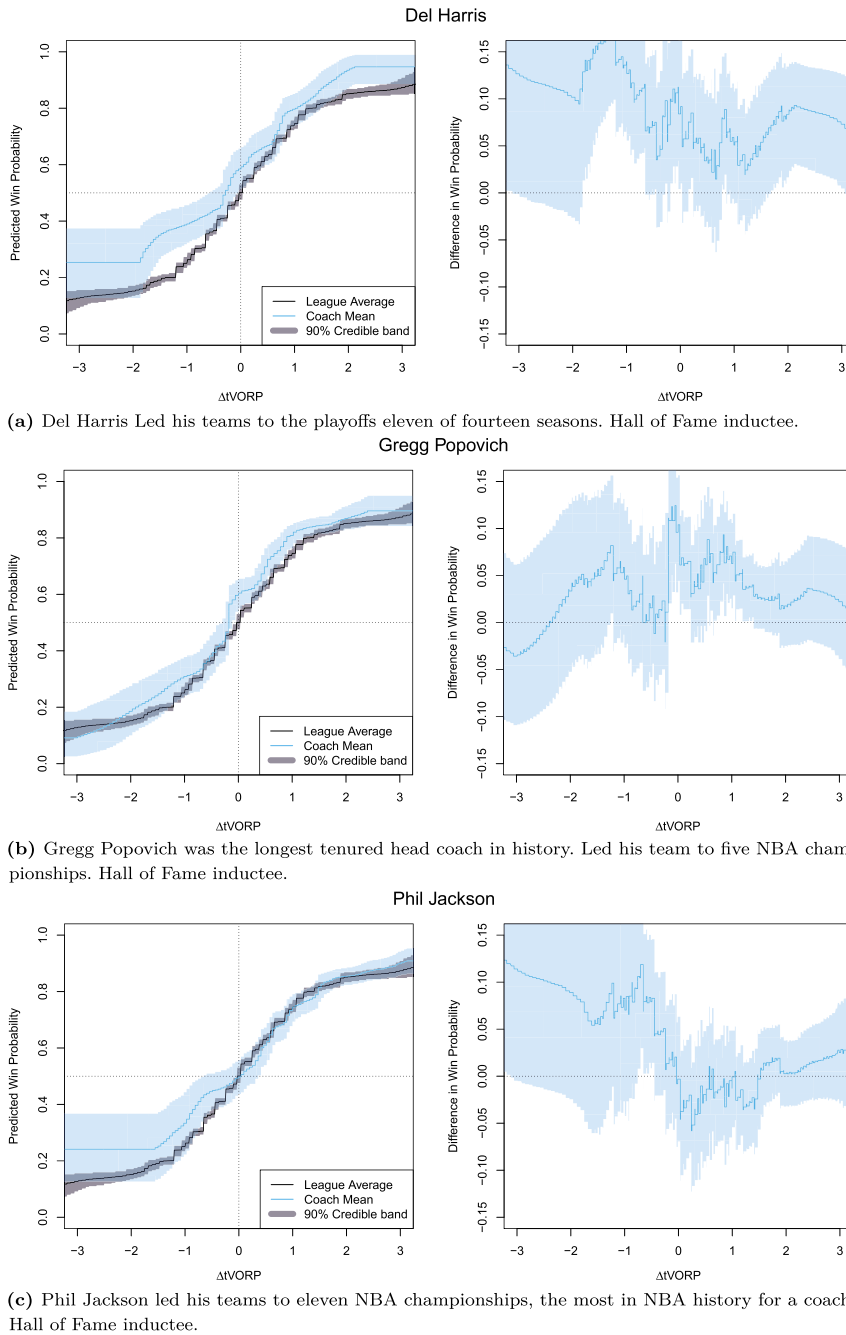
Figure 4: mBART posterior curves and 95 % credible bands for individual coaches compared with the "average" coach's on the left, the difference between the coach and "average" coach with 95 % credible bands on the right. Three Hall of Fame coaches represented.

of interest. The posterior mean and 95 % credible interval of $f_{\text{coach}}(\Delta\text{tVORP})$ for a few different coaches are shown in Figures 4–6. Each coach's curve is plotted against the "average" coach's curve. The posterior difference from "average" curve is also shown with a 95 % credible interval. The average mBART model is the fit on all of the available coaches' data. Three Hall of Fame coaches, three "average" coaches according to our metric, and three who are considered below-average coaches are represented. The three Hall of Fame coaches have values of ΔtVORP where

they have a statistically significantly higher win probability than average. We can see this visually with the three plots on the right side of Figure 4 where the credible band does not overlap the horizontal zero line. These represent values of ΔtVORP where we can say that the coach's scheme as we have defined it yields a higher chance to win than average. Winter (2010) reports that Tim Floyd, Kurt Rambis, and George Irvine are considered by some to be poor coaches and Figure 5 shows that their models reflect poorly on their scheme ability. We can see this visually with the three plots
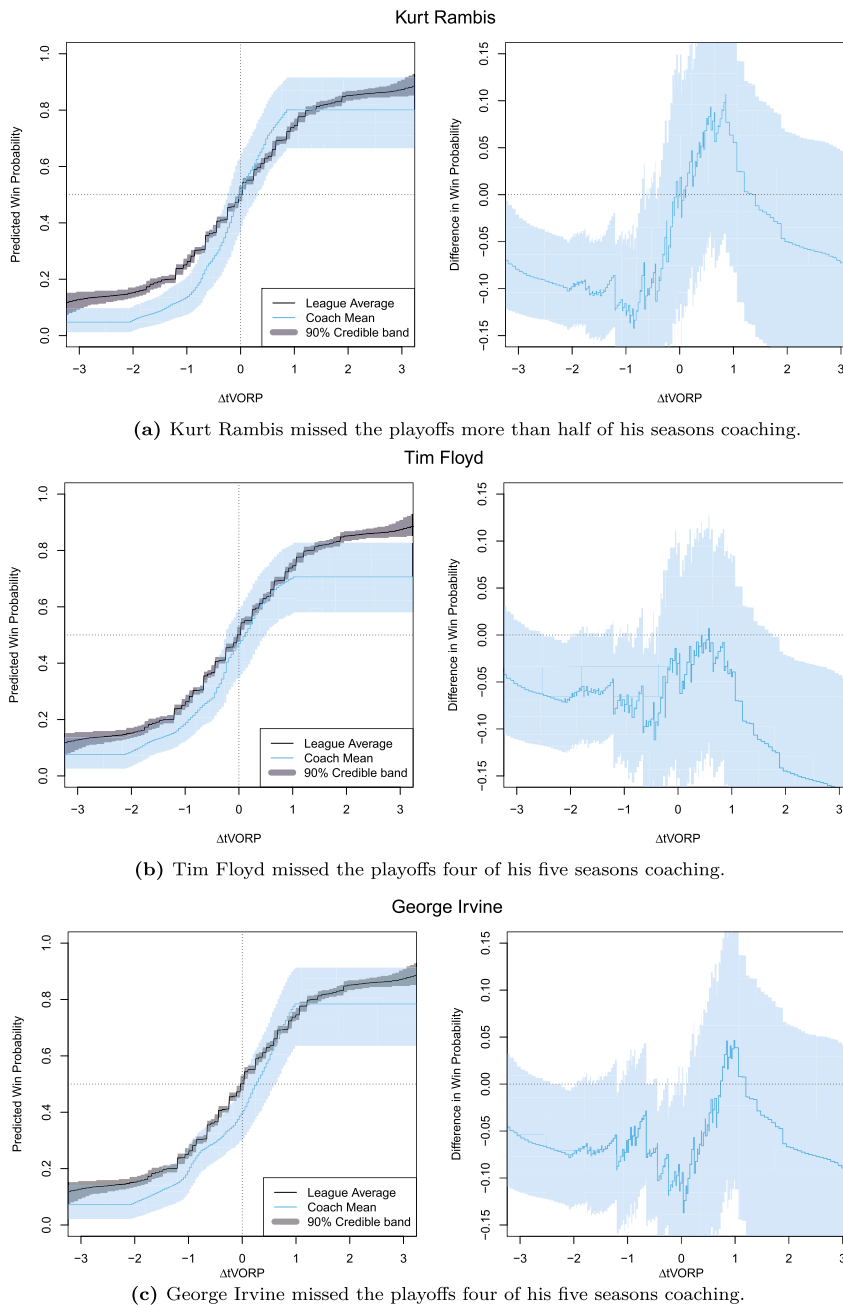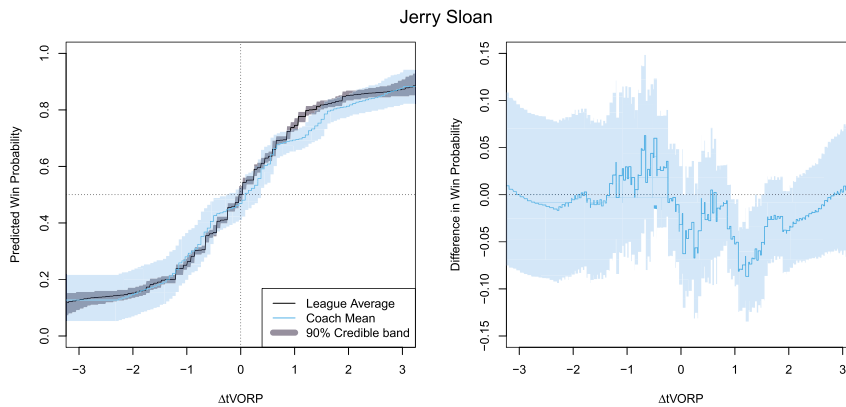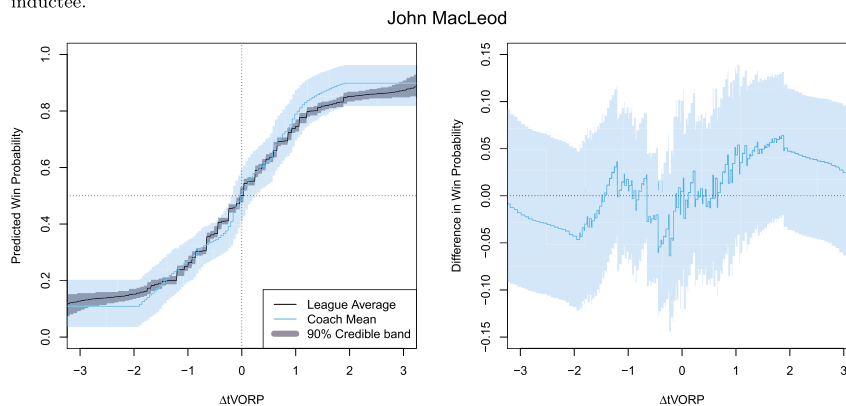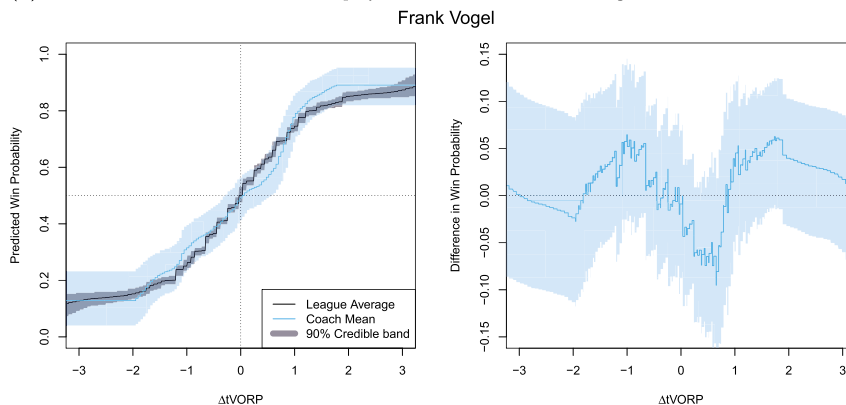

(a) Kurt Rambis missed the playoffs more than half of his seasons coaching.

(b) Tim Floyd missed the playoffs four of his five seasons coaching.

(c) George Irvine missed the playoffs four of his five seasons coaching.

Figure 5: mBART posterior curves and 95 % credible bands for individual coaches compared with the "average" coach's on the left, the difference between the coach and "average" coach with 95 % credible bands on the right. Three coaches with poor coaching records and shorter tenure represented.

(a) Jerry Sloan led his teams to the playoffs 20 of 23 seasons coaching the Utah Jazz. Hall of Fame inductee.

(b) John Macleod led his teams to the playoffs 11 of 18 seasons coaching.

(c) Frank Vogel led his teams to the playoffs eight of 12 seasons coaching.

Figure 6: mBART posterior curves and 95 % credible bands for individual coaches compared with the "average" coach's on the left, the difference between the coach and "average" coach with 95 % credible bands on the right. Three coaches who are not significantly different from the "average" coach very often, but had long coaching tenures and reached the playoffs most years.

on the right side of Figure 5 where the credible band does not overlap the horizontal zero line. These represent values of ΔtVORP where we can say that the coach's scheme as we have defined it yields a lower chance to win than average.

Figure 6 presents three coaches whose curves are fairly close to average. While the implication is that these coaches schemes were "average", we would like to highlight that many of the 30 teams in the NBA would like to have the

15th best coach. Furthermore, NBA coaching data is naturally a place where we see evidence of survivorship bias. When coaches lead their teams to good seasons they are retained. When coaches do not lead their teams to good seasons, particularly in the first few years of their tenure, they are not retained. For this reason, the coaches who are considered "poor" have less available data than those who are "average" or "great". This makes it more difficult to analyze coaches who had poor starts to their coaching

tenure. We see some evidence, however, that an "average" coach can be valuable to an NBA franchise. In Figure 6, we see three coaches who statistically are not very different from average but had very long tenures and some playoff success. For the most part, coaches do not need to separate themselves too far from average in terms of scheme ability to have long coaching careers in the NBA.

There are some interesting aspects of the model curves for a few of these coaches which are worth noting. Del Harris and Phil Jackson are notably better than average when their team's $\Delta tVORP$ was worse than their opponent. In other words when $\Delta tVORP < 0$, meaning the opposing team has a better on-court product than Harris or Jackson as measured by tVORP, they are winning a higher percentage of those games than average. Gregg Popovich has a large jump in win probability right around $\Delta tVORP = 0$, having a

projected win probability of nearly 60 %. This means when the opponent and Popovich's team have the same quality of court product, Popovich is projected to win six out of 10 of those games. Gaining an extra win out of every 10 over average in very close games can be worth a lot to a franchise.

Figure 7 shows density plots for 18 different coaches' projected win probability when $\Delta tVORP = 0$. When two teams' on-court product is equal, we would expect the baseline win probability to be 50 %. The coaches who are colored green in Figure 7 have a significantly higher projected win probability than 50 % and the coaches who are colored red in Figure 7 have a significantly lower projected win probability than 50 %. All others are colored blue and cannot be statistically separated from 50 % win probability. 85 % of all coaches represented in this analysis do not have a projected win probability at $\Delta tVORP = 0$ that is different than 50 %.
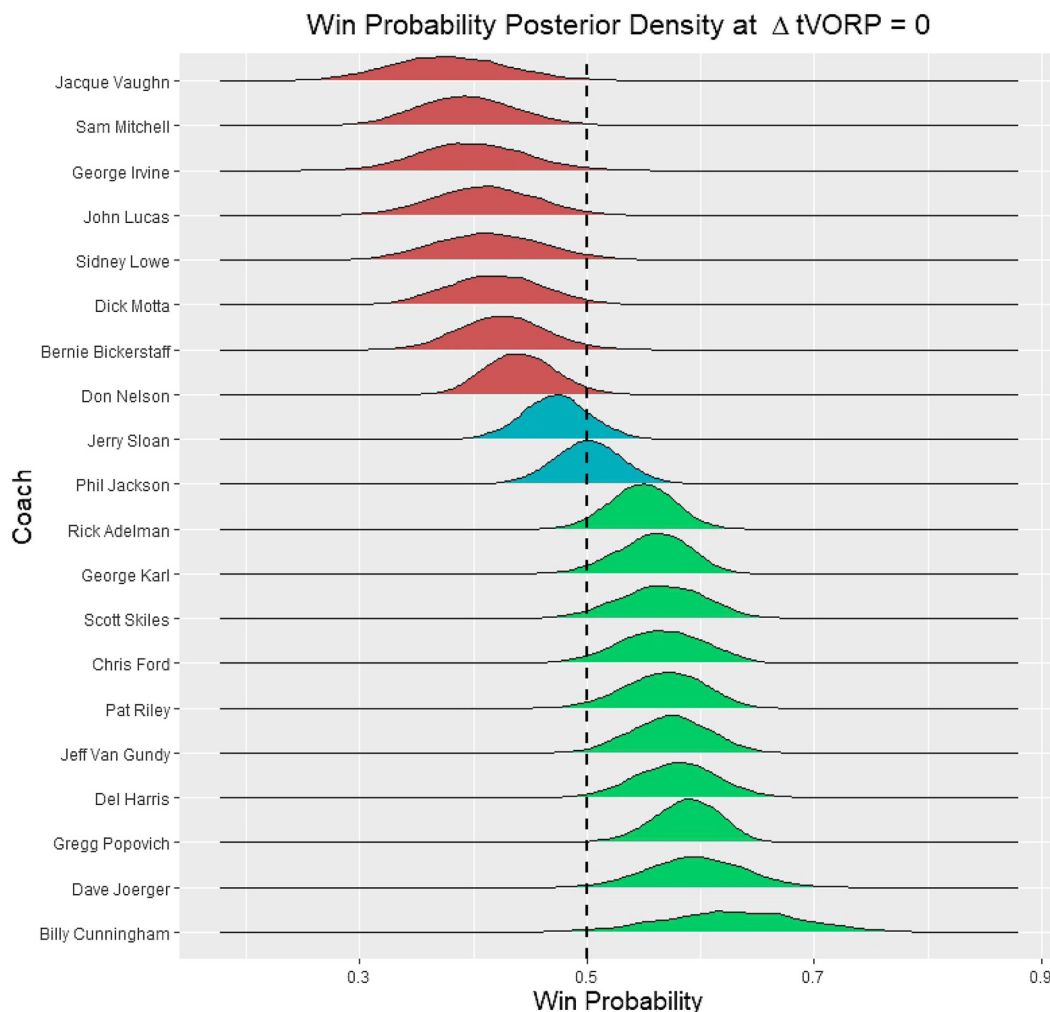

Figure 7: The density of the posterior draws for win probability for each coach at $\Delta tVORP = 0$. Red-colored densities indicate coaches with significantly lower than 50 % win probability, green-colored densities indicate coaches with significantly higher than 50 % win probability, and blue-colored coaches cannot be statistically separated from 50 % win probability. All coaches with posterior probability of at least 0.95 of being either higher or lower than 50 % are represented in this plot. For reference, we also include two coaches with posterior win probabilities that are not significantly different than 50 %. This illustrates that when team strength is equal, the vast majority of coaches do not have a predicted win probability different than 50 %.

Very few coaches have scheme ability which allows them to win more than half of their games where their on-court product is equal to their opponent.

Almost 40 % of NBA games in the dataset have ΔtVORP between -0.5 and 0.5. This means almost half of NBA games are played between opponents whose on-court product as we have defined it are very close to equal. If 40 % of games fall in this category then we are very interested in how coaches' scheme affects their team's win probability when on-court product is in this range. We can investigate expected win probability for coaches of interest using our fitted functions of ΔtVORP using Equation (6) in Section 3. To integrate out ΔtVORP and get the distribution of just the binary "Win" variable, we fit each coach's mBART with all observed ΔtVORP in the full data set as the test set. Calculating the average of all posterior draws for a set of values of ΔtVORP represents a coach's expected win probability using iterated expectations: $E(\text{win}) = E(E(\text{win}|\Delta\text{tVORP}))$. Table 2 in the Appendix shows the expected win probabilities for coaches over all values of ΔtVORP as well as when $\Delta\text{tVORP} \in (-0.5, 0.5)$. Included as well are the expected added wins over average in an 82 game season and for games where $\Delta\text{tVORP} \in (-0.5, 0.5)$ in an 82 game season.

By fitting $f_{\text{coach}}(\Delta\text{tVORP})$, we are able to approximate what coaches bring to the table beyond the strength of their roster and opposing roster. Many analysts have debated whether a coach like Phil Jackson won his titles because of his coaching ability or his roster strength. This analysis cannot definitively answer this question but it can help distinguish between coaches who contributed to winning with scheme and coaches who won more due to their roster strength. It is important to note that coaches also affect long-term playing development for their players if they work with them for a long time. They also manage personalities in the locker room which can help improve team chemistry. For example, Phil Jackson often had talented but challenging people on his roster, thus keeping everyone on the floor was a valuable contribution.

5 Discussion

The purpose of this analysis is to estimate the impact of NBA head coaches' in-game scheme on win probability when controlling for the players' skill and playing time. This is done by estimating the expected win probability $f_{\text{coach}}(\Delta\text{tVORP})$ using the new probit monotone Bayesian additive regression trees of (Fisher 2025). Players' skill and playing time is measured as the on-court product, as shown in Figure 1, and is approximated using ΔtVORP.

There are a few limitations to this analysis. First, tVORP can simply be mathematically optimized by playing the best players (as measured by highest VORP) for the full 48 min every game. In practice, this is not possible due to player fatigue, but tVORP could be altered in future analysis so it is optimized by playing the best players and lineups together for optimal time. Second, VORP is derived from BPM so it could suffer from similar drawbacks. BPM is limited in its ability to represent players' defensive impact since it is calculated using box scores which don't include advanced defensive metrics. Thus, tVORP may not be representing strong defensive players as accurately as offensive players.

Third, Figure 1 is only one of many possible diagrams that could be used. There are certainly variables one could argue contribute to on-court product or winning outside of the ones represented in this analysis. Thus, our analysis of saying one coach is better than average, or better than another, only holds for evaluating their "scheme" as we have defined it. However, by letting the data speak about the strength of coaching scheme after adjusting for players' ability, we are presenting an involved analysis of NBA coaching. We have to be careful when saying a coach is "better" or has better scheme ability than another. We are letting the data speak more clearly about the effects of scheme ability on winning and how it varies by coach using the variables we feel best proxy for coach scheme and team strength. We see in Figures 4–6 that the distribution of ΔtVORP is variable depending on the coach. We are skeptical about how well this represents coaches who never coached very good teams. Future analysis could investigate the best way to evaluate a coach who only ever coached poor teams.

Appendix: Supplementary tables

Table 2: Table of expected win probabilities and added wins over average for each coach. Included is the expected win probability for all values of ΔtVORP, expected added wins over average for an 82 game season, expected win probability when ΔtVORP $\in$ (−0.5, 0.5), and expected added wins over average for games in an 82 game season where ΔtVORP $\in$ (−0.5, 0.5), usually around 33 games.

Coach	All games		Games with ΔtVORP $\in$ (−0.5, 0.5)	
	E(Win)	Added wins	E(Win)	Added wins
Billy Cunningham	0.60	8.40	0.61	3.50
Dave Joerger	0.60	8.40	0.59	3.10
Gene Shue	0.54	3.40	0.58	2.70
Ron Rothstein	0.54	3.20	0.58	2.60
Del Harris	0.60	7.90	0.57	2.40
K.C. Jones	0.55	4.20	0.56	2.10
Jeff Van Gundy	0.55	3.70	0.56	2.00
Paul Westhead	0.51	0.70	0.56	2.00
Chris Ford	0.54	3.30	0.56	1.90
Gregg Popovich	0.56	5.30	0.56	1.80
Pat Riley	0.56	4.70	0.55	1.80
Brad Stevens	0.55	3.80	0.55	1.80
Steve Kerr	0.57	5.90	0.55	1.70
Lawrence Frank	0.52	1.70	0.55	1.70
Larry Bird	0.60	8.10	0.55	1.60
Garry St. Jean	0.53	2.40	0.55	1.50
Scott Skiles	0.52	1.60	0.54	1.40
Bill Fitch	0.53	2.10	0.54	1.30
Jay Triano	0.54	3.40	0.54	1.30
Rick Adelman	0.53	2.60	0.54	1.30
Richie Adubato	0.55	3.90	0.54	1.30
Mike Brown	0.53	2.30	0.54	1.30
Maurice Cheeks	0.53	2.40	0.54	1.20
Phil Johnson	0.49	−0.70	0.54	1.20
Rudy Tomjanovich	0.55	3.80	0.54	1.20
Chuck Daly	0.55	4.00	0.54	1.20
George Karl	0.54	3.20	0.53	1.00
Luke Walton	0.53	2.20	0.53	1.00
Wes Unseld	0.56	4.90	0.53	1.00
Jason Kidd	0.55	3.90	0.53	0.90
Frank Layden	0.53	2.40	0.53	0.90
Bob Weiss	0.53	2.20	0.53	0.90
Rick Carlisle	0.56	4.80	0.53	0.80
Kevin Loughery	0.53	2.50	0.52	0.70
Paul Westphal	0.54	3.10	0.52	0.60
Michael Malone	0.52	1.50	0.52	0.60
Mike Dunleavy	0.53	2.20	0.52	0.60
Vinny Del Negro	0.53	2.40	0.52	0.60
Mike Budenholzer	0.53	2.60	0.52	0.60
Stan Albeck	0.54	3.00	0.52	0.60
Willis Reed	0.50	−0.00	0.52	0.50
Brett Brown	0.50	−0.20	0.52	0.50
Doug Moe	0.51	1.20	0.52	0.50
Isiah Thomas	0.54	3.60	0.52	0.50
Rick Pitino	0.50	0.30	0.51	0.50
Stan Van Gundy	0.53	2.40	0.51	0.50
Steve Clifford	0.51	0.90	0.51	0.50
Jim O'Brien	0.50	−0.00	0.51	0.40

Table 2: (continued)

Coach	All games		Games with $\Delta tVORP \in (-0.5, 0.5)$	
	E(Win)	Added wins	E(Win)	Added wins
Dave Cowens	0.54	3.10	0.51	0.40
Erik Spoelstra	0.53	2.50	0.51	0.30
Mike Fratello	0.52	1.70	0.51	0.30
Monty Williams	0.49	−0.60	0.51	0.30
Kevin McHale	0.51	1.20	0.51	0.30
Don Casey	0.49	−0.80	0.51	0.30
Terry Porter	0.53	2.10	0.51	0.30
Cotton Fitzsimmons	0.51	0.40	0.51	0.30
Avery Johnson	0.52	1.90	0.51	0.20
Kenny Atkinson	0.51	0.70	0.51	0.20
Dwane Casey	0.53	2.20	0.51	0.20
Phil Jackson	0.54	3.20	0.51	0.20
Tyrone Corbin	0.55	4.10	0.51	0.20
P.J. Carlesimo	0.52	1.90	0.51	0.20
Nick Nurse	0.59	7.60	0.50	0.10
Tom Thibodeau	0.51	1.20	0.50	0.10
Doc Rivers	0.51	1.00	0.50	0.10
Flip Saunders	0.51	1.00	0.50	0.00
Larry Brown	0.53	2.20	0.50	−0.00
Nate McMillan	0.53	2.70	0.50	−0.10
Jeff Hornacek	0.52	1.30	0.49	−0.20
Kurt Rambis	0.49	−0.50	0.49	−0.20
Danny Ainge	0.53	2.70	0.49	−0.20
Hubie Brown	0.50	0.20	0.49	−0.20
Alvin Gentry	0.51	0.50	0.49	−0.30
Jerry Sloan	0.50	0.40	0.49	−0.30
Larry Drew	0.51	1.10	0.49	−0.30
John MacLeod	0.52	1.90	0.49	−0.30
Johnny Davis	0.49	−1.20	0.49	−0.30
Mike D'Antoni	0.50	−0.10	0.49	−0.40
Jack Ramsay	0.51	1.10	0.49	−0.40
Keith Smart	0.51	0.80	0.49	−0.40
Bill Russell	0.45	−4.20	0.49	−0.50
Don Chaney	0.50	−0.20	0.49	−0.50
Byron Scott	0.50	−0.40	0.49	−0.50
Lenny Wilkens	0.51	1.10	0.48	−0.50
Doug Collins	0.48	−1.80	0.48	−0.50
Fred Hoiberg	0.51	0.40	0.48	−0.50
Mark Jackson	0.50	0.10	0.48	−0.60
Frank Vogel	0.52	1.60	0.48	−0.60
Tyronn Lue	0.50	0.20	0.48	−0.70
Johnny Bach	0.44	−4.60	0.48	−0.70
Randy Wittman	0.50	0.00	0.47	−0.90
David Fizdale	0.47	−2.80	0.47	−0.90
Eddie Jordan	0.49	−1.10	0.47	−0.90
Lionel Hollins	0.49	−0.50	0.47	−0.90
Matt Guokas	0.50	0.40	0.47	−1.00
Allan Bristow	0.52	1.60	0.47	−1.00
Terry Stotts	0.50	−0.40	0.47	−1.00
Scott Brooks	0.48	−1.60	0.47	−1.00
Mike Schuler	0.50	−0.20	0.47	−1.00
Dan Issel	0.51	0.90	0.47	−1.00
Eric Musselman	0.50	−0.30	0.47	−1.10
Dick Motta	0.50	−0.10	0.46	−1.20

Table 2: (continued)

Coach	All games		Games with ΔtVORP $\in (-0.5, 0.5)$	
	E(Win)	Added wins	E(Win)	Added wins
Tim Floyd	0.46	−3.30	0.46	−1.20
Mike Woodson	0.48	−1.40	0.46	−1.30
Paul Silas	0.49	−0.90	0.46	−1.30
Bob Hill	0.48	−1.20	0.46	−1.30
Don Nelson	0.52	1.30	0.46	−1.40
Quin Snyder	0.48	−1.40	0.46	−1.40
Jim Lynam	0.48	−1.30	0.45	−1.60
Bernie Bickerstaff	0.50	0.30	0.45	−1.60
J.B. Bickerstaff	0.47	−2.40	0.45	−1.70
Bill Musselman	0.44	−5.20	0.45	−1.80
Billy Donovan	0.46	−2.90	0.45	−1.80
Brian Hill	0.47	−2.30	0.45	−1.80
Sam Mitchell	0.48	−1.70	0.43	−2.20
Sidney Lowe	0.44	−5.10	0.43	−2.40
Jacque Vaughn	0.41	−7.70	0.43	−2.40
Jimmy Rodgers	0.49	−1.10	0.42	−2.50
George Irvine	0.46	−3.00	0.42	−2.70
John Lucas	0.43	−5.80	0.40	−3.20

References

Berry, C. R. and Fowler, A. (2019). How much do coaches matter? In: *Sports analytic conference*. MIT Sloan, Boston, MA.

Chen, Y., Dai, J., and Zhang, C. (2019). A neural network model of the NBA most valued player selection prediction. In: *Proceedings of the 2019 the international conference on pattern recognition and artificial intelligence*, pp. 16−20.

Chipman, H. A., George, E. I., and McCulloch, R. E. (1998). Bayesian CART model search. *J. Am. Stat. Assoc.* 93: 935−948.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: bayesian additive regression trees. *Ann. Appl. Stat.* 4: 266−298.

Chipman, H. A., George, E. I., McCulloch, R. E., and Shively, T. S. (2022). mBART: multidimensional monotone BART. *Bayesian Anal.* 17: 515−544.

Dawson, P., Dobson, S., and Gerrard, B. (2000). Estimating coaching efficiency in professional team sports: evidence from English association football. *Scot. J. Polit. Econ.* 47: 399−421.

Engelmann, J. (2014). The 'Effect of Being Up X'. *Blog*, Available at: https://apbr.org/metrics/viewtopic.php?f=2&t=8501.

Fisher, J. D. (2025). Probit monotone BART. *arXiv*. 2509.00263.

Fort, R., Lee, Y. H., and Berri, D. (2008). Race, technical efficiency, and retention: the case of NBA coaches. *Int. J. Sport Finance* 3: 84.

Goldman, M. and Rao, J. M. (2013). Live by the Three, Die by the Three? The Price of Risk in the NBA. In: *Submission to the MIT Sloan sports analytics conference*, p. 155.

Hofler, R. A. and Payne, J. E. (2006). Efficiency in the National Basketball Association: a stochastic Frontier approach with panel data. *Manag. Decis. Econ.* 27: 279−285.

McCorey, J. M. (2021). Forecasting Most valuable players of the National Basketball Association, PhD thesis. The University of North Carolina at Charlotte.

Myers, D. (2020). About Box Plus/Minus. *Blog*, Available at: https://www.basketball-reference.com/about/bpm2.html.

Winter, J. (2010). NBA power rankings: the 50 worst coaches in NBA history. *Blog*, Available at: https://bleacherreport.com/articles/514524-nba-power-rankings-the-50-worst-coaches-in-nba-history.

Woolner, K. (2002). Understanding and measuring replacement level. *Baseb. Prospectus* 1: 55−66.

Wu, W., Feng, K., Li, R., Sengupta, K., and Cheng, A. (2018). Classification of NBA salaries through player statistics. *Sports Anal. Group Berk.*, https://sportsanalytics.berkeley.edu/projects/nba-salaries-stats.pdf (Accessed 19 September 2020).