Winning Advantage and Its Applications*

Kyle Y. Lin† Ken K. Lin‡

January 1, 2007

Abstract

In most team sports, the whole game can be broken up into a sequence of plays, each of which involves a subset of players and contributes differently to the game's outcome. We develop an innovative concept—called the winning advantage—to quantify the value of each play, and use it to evaluate each player's contribution toward winning or losing a game. Winning advantage evaluates all plays on the same scale, so that contributions from all players in different types of plays can be directly compared. An example in baseball is provided to demonstrate the winning advantage and its applications.

Keywords: winning advantage, winning probability, sport, baseball, evaluation, Markov chain.

1 Introduction

In a sport game, the state of the game changes over time as the game progresses. From a team's standpoint, each state possesses an intrinsic winning advantage—the chance of winning that is attributed to the state. Each play moves the game from one state to another, and the numerical difference in winning advantage between the two states quantifies the value of the play. In addition, by properly crediting this difference in winning advantage to responsible players and accumulating each player's contribution throughout the game, we can determine each player's contribution toward winning or losing the game.

Generally speaking, a team enjoys a higher winning advantage if it is in a better position to win. In a stochastic game where the probability measure of all random events is well defined so that each side can play with the optimal strategy, the winning advantage coincides with the winning probability (such as Blackjack). In sports, however, the winning probability depends largely on the abilities of the players, and there does not exist a proper probability

*The authors thank Retrosheet for providing the play-by-play Major League Baseball data. More results are available at www.whoismvp.com.

†1411 Cunningham Road, Operations Research Department, Naval Postgraduate School, Monterey, California 93943-5219. E-mail: kylin@nps.edu.

‡E-mail: ken.lin@stanfordalumni.org.

measure for all random events, so it is impossible to define the winning probability. The idea of the winning advantage is to quantify the chance of winning that is attributed to the state of the game, by excluding the chance of winning that is attributed to the players' abilities.

The definition of a state varies in different sports. In baseball, the state includes the top or the bottom of the inning, the difference in runs scored, the number of outs, and the base-occupying situation. In basketball, the state includes the time remaining, the possession of the ball, and the difference in scores. In football, the state includes the time remaining, the possession and location of the ball, the down, the number of yards needed for a first down, and the difference in scores. Although states are defined differently, using a Markov chain model allows us to develop a unified method to calculate the winning advantage in different sports.

Using the winning advantage to evaluate each player's performance has three distinctive features compared with traditional statistics in sports:

1. The winning advantage evaluates all plays on the same scale, so that all players' contributions can be directly compared—even across sports.

2. The contribution in winning advantage can be directly interpreted as the number of games a player helps his team win.

3. The winning advantage rewards great performances in clutch situations. For example, a home run in the top of the ninth inning when the scores are tied contributes much more to the winning than a home run in the top of the first inning.

1.1 Relation to Existing Work

In sports, there are numerous statistics that evaluate a player's performance. The most popular statistics remain to be those calculated by counting (such as home runs hit in baseball and yards rushed in football) and by taking averages (such as batting average in baseball and points per game in basketball). Each of these statistics, however, measures one dimension of the game, so academic researchers have proposed many methods to integrate these one-dimensional statistics into one statistic that measures the overall productivity of a player.

For example in baseball, Cover and Keilers (1977) builds a model to predict the number of runs a team would score if one player bats in all nine positions in the lineup. Pankin (1978) evaluates the offensive performance by the increase in expected runs by the batter. A few other methods can be found in Bennett and Flueck (1983). These methods evaluate a player by how many runs a player helps his team score, because runs are closely tied to wins. Our method, however, takes a step further and evaluates a player directly by how many games a player helps his team win. In basketball, statistical regression models that are used to measure the production of basketball players can be found in Zak, Huang, and Siegfried (1979), Hofler and Payne (1997), and Berri (1999). These models are different from ours because they assemble the production statistic by taking weighted average over one-dimensional statistics, whereas our model directly evaluates how much each play contributes to the outcome of a game.

The works that are most closely related to our approach are those reported in Mills and Mills (1970), Bennett and Flueck (1984), and Bennett (1993). In these works, the authors first estimate the *winning probability* for each state in baseball, and then compute statistics based on these probabilities. However, their methods of calculating the winning probability are restricted to baseball. In addition, their calculation relies on many simplifying assumptions and ignores the asymmetry in baseball—the away team batting on the top and the home team the bottom of each inning. The strength of our method of estimating the winning advantage is the use of a Markov chain model. Such a Markov chain model naturally reflects the asymmetry in baseball and can be used directly in other sports.

1.2 Overview and Outline

In this paper, we use a Markov chain model to describe the evolution of a sport game. With each state properly defined in a sport, we present an algorithm to recursively compute the winning advantage of each state. The algorithm uses conditional expectation to improve estimation, and applies to any sport as long as one can construct an appropriate Markov chain model. Once we find the winning advantage for each state, we can credit each play with the winning advantage difference resulting from state transition. Players involved in a play can be properly credited, and their contribution toward the outcome of a game can be determined. We use an example in baseball to demonstrate our method.

The rest of this paper is organized as follows. In Section 2, we define the winning advantage and present an algorithm to compute it. In Section 3, we explain how to use the winning advantage to score each play and credit players involved in the play. Finally, we offer some concluding remarks in Section 4.

2 Calculation of the Winning Advantage

The *winning advantage* of a state represents the chance of winning that is attributed to the state itself and does not include the chance of winning that is attributed to the players. For example, leading by 2 runs gives a team a better chance to win than leading by 1 run, so the former state—when everything else held equal—has a higher winning advantage than the latter. We arbitrarily choose to define the winning advantage from the home team's standpoint for the ease of exposition.

Let Ω denote the set of all states in a game, and for $s \in \Omega$, let $w(s)$ denote the winning advantage for state s. Because $w(s)$ is the winning probability if we randomly select two teams to start playing in state s, it is a function of s and does not depend on the history of the game before state s is reached. In other words, we assume that the evolution of the game can be described by a Markov chain.

Suppose we are given a collection of data with the detailed play-by-play descriptions of n games. The easiest way to estimate $w(s)$, $s \in \Omega$, is to first let n_s denote the number of games that have reached state s at some point, and observe that each of these n_s games would be independently won by the home team with probability $w(s)$. Letting $n_{s,w}$ denote the number of times the home team does win in those n_s games, the maximum likelihood estimator (see page 218 in Ross (2000) for instance) for $w(s)$ is

$$\frac{n_{s,w}}{n_s}. \tag{1}$$

For example, consider the data set that includes 65949 regular-season games played in Major League Baseball between 1974 and 2005.[1] The home team won 35593 times in 65949 games, so from Equation (1) the winning advantage for the home team in the beginning of the game is 0.5397. Out of those 65949 games, a leadoff home run occurred 975 times, and the home team won 438 times in those games, so the winning advantage becomes 0.4492 if the leadoff hitter of the opposing team hit a home run. Table 1 gives the winning advantage calculated by Equation (1) for states after the leadoff hitter's plate appearance is terminated.

Table 1: The Effect of the leadoff hitter; data from Major League 1974–2005.

Outcome of the leadoff hitter	Number of wins	Number of occurrences	Winning advantage by Equation (1)
Out	24875	44063	0.5645
Reach 1st	8628	17374	0.4966
Reach 2nd	1369	2916	0.4695
Reach 3rd	283	621	0.4557
Home Run	438	975	0.4492
Overall	35593	65949	0.5397

The drawback of using Equation (1) to compute the winning advantage is that the sample size can be small if state s rarely appears in the data set. To improve the estimator in Equation (1), let Ω_s denote the set of states that can be directly reached from state s. From our data set that consists of n games, let $n_{s,t}$ denote the number of games that move from state s directly into state t, and $n_{s,t,w}$ the number of games that move from state s directly into state t and are eventually won by the home team. We can rewrite the estimator in Equation (1) as

$$\frac{n_{s,w}}{n_s} = \sum_{t\in\Omega_s} \frac{n_{s,t}}{n_s}\frac{n_{s,t,w}}{n_{s,t}}.$$

Note that $n_{s,t,w}/n_{s,t}$ is an unbiased estimator for $w(t)$, but a better unbiased estimator for $w(t)$ is $n_{t,w}/n_t$, because the latter has a larger sample size. Therefore, a better estimator for $w(s)$ is

$$\sum_{t\in\Omega_s} \frac{n_{s,t}}{n_s}\frac{n_{t,w}}{n_t}. \tag{2}$$

Another way to see why the preceding estimator is better than the estimator $n_{s,w}/n_s$ is to express $w(s)$ by conditioning on what happens next in state s:

$$w(s) = \sum_{t\in\Omega_s} p_{s,t} w(t),$$

[1] Play-by-play data is provided by Retrosheet. Games played in the 1999 season are not included because data is unavailable. Games that are called or tied are not included.

where $p_{s,t}$ denotes the probability that the Markov chain, when in state s, will next visit state t. In other words, in Equation (2) we use $n_{s,t}/n_s$ to estimate $p_{s,t}$ and use $n_{t,w}/n_t$ to estimate $w(t)$.

To take this idea one step further, once we have an estimator for $w(t)$, say $\tilde{w}(t)$, for all $t \in \Omega_s$, we can estimate $w(s)$ by

$$\tilde{w}(s) = \sum_{t \in \Omega_s} \frac{n_{s,t}}{n_s} \tilde{w}(t). \tag{3}$$

Because the game always moves forward and never visits the same state more than once, there does not exist a loop in the state space. Therefore, we can use Equation (3) to recursively estimate the winning advantage backward for all states.

We next demonstrate how to use Equation (3) to estimate the winning advantage with a baseball example. We use a vector (i, j, k, l, m) to denote the state of a baseball game, where $i \in \{1, 2, \ldots\}$, $j \in \{t, b\}$, $k \in \{-32, -31 \ldots, -1, 0, 1, \ldots, 31, 32\}$, $l \in \{0, 1, 2\}$, and $m \in \{000, 001, 010, \ldots, 111\}$. The argument i represents the inning; j indicates whether it is the top or the bottom of the inning; k the margin of leading for the home team; l the number of outs; m the base-occupying situation. For the base-occupying situations, we let $m = 000$ when the bases are empty; $m = 001$ when there is a man on first; $m = 011$ if there are men on first and second; $m = 111$ if bases are loaded; and so on. In the state space, we allow the run difference to be at most 32, which is consistent with our data set where the largest difference in runs is 22.

To use Equation (3), we first need to have an estimator for $w(10, t, 0, 0, 000)$—the winning advantage when the game goes to extra innings. Because in our data set there are 6077 extra-inning games, and the home team won 3170 times, we let $\tilde{w}(10, t, 0, 0, 000) = 3170/6077$. We next present the algorithm based on Equation (3); some explanations follow.

Algorithm

1. Set $\tilde{w}(10, t, 0, 0, 000) = 3170/6077$.

2. Set $\tilde{w}(9, b, k, l, m) = 1$ for $k \geq 1$, $l = 0, 1, 2$, and $m = 000, 001, \ldots, 111$.

3. Set $\tilde{w}(i, t, -32, l, m) = 0$ for $i = 1, 2, \ldots, 9$, $l = 0, 1, 2$, and $m = 000, 001, \ldots, 111$.

4. Set $\tilde{w}(i, b, 32, l, m) = 1$ for $i = 1, 2, \ldots, 8$, $l = 0, 1, 2$, and $m = 000, 001, \ldots, 111$.

5. Use the following loop to compute the winning advantage:

```
for i ← 9 to 1
   j ← b
   for k ← 31 to −32          (if i = 9, change 31 to 0)
      for l ← 2 to 0
         for m ← 7(111) to 0(000)
            compute w̃(i, j, k, l, m) according to Equation (3).
```

$j \leftarrow t$
for $k \leftarrow -31$ to 32
for $l \leftarrow 2$ to 0
for $m \leftarrow 7(111)$ to $0(000)$
compute $\tilde{w}(i, j, k, l, m)$ according to Equation (3).

6. Set $\tilde{w}(i, j, k, l, m) = \tilde{w}(9, j, k, l, m)$, for $i = 10, 11, \ldots$, and for all j, k, l, m.

In step 1 we set the winning advantage for an extra-inning game; in step 2 we set the winning advantage to 1 if the home team wins at the bottom of the ninth inning. In steps 3 and 4 we assign 1 to the winning advantage if a team leads by 32 runs (or more). In other words, we assume it is impossible to overcome a deficit of 32 runs in baseball, which is consistent with our data set. In step 5, we use Equation (3) to estimate the winning advantage from the bottom of the ninth inning to the top of the first inning. Finally in step 6 we assign the winning advantage in extra innings equal to its counterpart in the ninth inning. The rationale is that in each extra inning, either one team scores more runs to win the game, or they have to play the next extra inning. Therefore, the situation in each extra inning is just like the situation in the ninth inning, so the winning advantage in state $(9, j, k, l, m)$ is equal to that in state (i, j, k, l, m) for $i = 10, 11, \ldots$, and for all j, k, l, m. Table 2 gives a part of our results with this estimation method.

Table 2: Winning advantage in the bottom of the first inning; data from Major League Baseball 1974–2005.

		m							
k[a]	l	111	110	101	100	011	010	001	000
1	0	0.8181	0.8008	0.7905	0.7681	0.7676	0.7461	0.7284	0.6950
1	1	0.7584	0.7586	0.7395	0.7334	0.7233	0.7071	0.6940	0.6730
1	2	0.6992	0.6888	0.6852	0.6787	0.6805	0.6727	0.6643	0.6537
0	0	0.7516	0.7311	0.7117	0.6801	0.6884	0.6547	0.6327	0.5916
0	1	0.6896	0.6743	0.6548	0.6422	0.6300	0.6090	0.5949	0.5647
0	2	0.6116	0.5912	0.5873	0.5753	0.5813	0.5695	0.5595	0.5454
−1	0	0.6688	0.6431	0.6240	0.5800	0.5957	0.5589	0.5333	0.4876
−1	1	0.5914	0.5757	0.5561	0.5403	0.5323	0.5072	0.4903	0.4575
−1	2	0.5070	0.4768	0.4837	0.4636	0.4728	0.4610	0.4530	0.4361

[a] k: run difference; l: outs; m: base-occupying situation.

3 Play-by-Play Evaluation

When a play moves the game from state s to state t, we say the *winning advantage score* (WAS) of the play is $w(t) - w(s)$, which is then credited to the responsible player. If two or more players are involved in the play, then depending on the nature of a sport, one can decide how to divide WAS to responsible players. We next use an example in baseball to demonstrate this play-by-play evaluation.

In Game 1 of the 2004 American League Championship Series, New York Yankees were the home team, and the Boston Red Sox were the away team. As shown in Table 3, when the game started, the state was $(1, t, 0, 0, 000)$, whose winning advantage is 0.5397 (from home team's standpoint). The strikeout of the leadoff hitter moved the state to $(1, t, 0, 1, 000)$, whose winning advantage is 0.5632. Therefore, from this strikeout Mussina earned 0.0235 WAS, and Damon lost 0.0235 WAS. The second play of the game was a fly out by Bellhorn, and the state became $(1, t, 0, 2, 000)$, whose winning advantage is 0.5801. Mussina earned 0.0169 WAS by inducing this fly out, and Bellhorn lost 0.0169 WAS. The game continued and so did the play-by-play evaluation. In the last play of the game, Mueller grounded into double play to reach the state $(9, b, 3, 0, 000)$; the New York won the game by 3 runs, and the winning advantage became 1.

Table 3: 2004 American League Championship Series Game 1.

Inning	Play	Resulting state	WA	WAS	BOS player	NYY player
New game	play ball	$1, t, 0, 0, 000$	0.5397			
1 Top	strike out	$1, t, 0, 1, 000$	0.5632	0.0235	Damon	Mussina
	fly out	$1, t, 0, 2, 000$	0.5801	0.0169	Bellhorn	Mussina
	ground out	$1, b, 0, 0, 000$	0.5916	0.0115	Ramirez	Mussina
1 Bot	fly out	$1, b, 0, 1, 000$	0.5647	−0.0269	Schilling	Jeter
	fly out	$1, b, 0, 2, 000$	0.5454	−0.0193	Schilling	Rodriguez
	double	$1, b, 0, 2, 010$	0.5695	0.0241	Schilling	Sheffield
	double	$1, b, 1, 2, 010$	0.6727	0.1032	Schilling	Matsui
	single	$1, b, 2, 2, 001$	0.7566	0.0839	Schilling	Williams
	ground out	$2, t, 2, 0, 000$	0.7399	−0.0167	Schilling	Posada
⋮	⋮	⋮	⋮	⋮	⋮	⋮
		$9, t, 3, 0, 000$	0.9778			
9 Top	fly out	$9, t, 3, 1, 000$	0.9912	0.0134	Nixon	Rivera
	single	$9, t, 3, 1, 001$	0.9766	−0.0146	Varitek	Rivera
	single	$9, t, 3, 1, 011$	0.9395	−0.0371	Cabrera	Rivera
	double play	$9, b, 3, 0, 000$	1.0000	0.0605	Mueller	Rivera

Throughout a game, the WAS accumulated by each player tells how much he contributes toward the outcome of the game. For example, in Game 1 of the 2004 American League Championship Series, Matsui of the New York Yankees was involved in five plays with respective WAS 0.1032, 0.0611, −0.0017, 0.0021, and −0.0209, so his personal contribution for this game is 0.1438. Furthermore, because the WAS is additive, we can evaluate a player's contribution by accumulating his WAS in a post-season series, in a regular season, or in his entire career. Table 4 presents WAS numbers for the top ten pitchers in 2005 Major League regular-season games, and Table 5 presents those for the top ten position players.

Table 4: Major League Baseball top ten pitchers in 2005.

Pitcher	Team	Games	IP[a]	H	R	BB	SO	WAS
Roger Clemens	Hou	32	211.1	151	50	62	185	6.5932
Dontrelle Willis	Fla	40	236.1	213	71	55	170	6.5776
Andy Pettitte	Hou	33	222.1	188	64	41	171	5.4943
Jake Peavy	SD	30	203.0	162	68	50	216	5.1220
Chris Carpenter	StL	33	241.2	204	76	51	213	4.6004
Roy Halladay	Tor	19	141.2	118	38	18	108	4.5189
Roy Oswalt	Hou	35	241.2	243	84	48	184	4.3940
Johan Santana	Min	33	231.2	180	75	45	238	4.2263
Carlos Zambrano	ChC	33	223.1	170	85	86	202	4.0604
Huston Street	Oak	67	78.1	53	26	26	72	3.6240

[a]IP: innings pitched; H: hits; R: runs allowed; BB: walks; SO: strike outs; WAS: winning advantage score.

We can also use WAS to interpret the number of games a player helps his team win. To understand the calculation, first consider a sport in which the winning advantage in the beginning of a game is 0.5 for either side. In this sport, the players on the winning team accumulate $1 - 0.5 = 0.5$ in WAS, and those on the losing team accumulate -0.5. In other words, each 0.5 in WAS corresponds to a win, and each -0.5 corresponds to a loss. Therefore, 2 times WAS becomes the number of games a player wins for his team. Although in some sports, baseball for instance, the winning advantage in the beginning of a game is slightly different from 0.5, this calculation is still reasonable, because each team plays the same number of games at home and on the road. For example, as shown in Tables 4, Roger Clemens contributed 13.19 wins by his pitching performance in 2005.

Because the WAS is directly tied to winning, the player who accumulates the highest WAS should be the strong favorite for the most valuable player award. This method can apply to a single game, a short series, or an entire season. The significance of using WAS to select the most valuable player is that all players—regardless their respective positions—can be directly compared on the same scale.

In baseball, it is relatively easy to find the responsible player for most plays because each play can be viewed as a discrete event independent of others. When more than one player is involved, the WAS can be divided among them based on their contribution (or responsibility). In some other sports, however, it may not be straightforward to allocate WAS. For example in basketball, the game progresses continuously unless a foul is called or a timeout is taken. Many plays—such as a fast break or a turnover—involve more than one player, and it is not clear who contributes the most. One way to credit the players during a continuous play is to divide WAS equally among all five players on the court. A player would receive positive WAS if his team moves to a better state during the time he is playing.

Besides analyzing a player's contribution, we can also use WAS to analyze a team's

Table 5: Major League Baseball top ten position players in 2005.

Batter	Team	Games	PA[a]	HR	RBI	BA	OBP	SLG	WAS
David Ortiz	Bos	159	713	47	148	0.2995	0.4039	0.6040	8.9771
Alex Rodriguez	NYY	162	715	48	130	0.3207	0.4308	0.6099	6.1571
Chipper Jones	Atl	109	432	21	72	0.2961	0.4213	0.5559	5.5484
Carlos Delgado	Fla	144	616	33	115	0.3013	0.4042	0.5816	5.2777
Tony Clark	Ari	130	393	30	87	0.3037	0.3715	0.6361	4.9836
Travis Hafner	Cle	137	578	33	108	0.3045	0.4170	0.5947	4.9394
Derrek Lee	ChC	158	691	46	107	0.3350	0.4284	0.6616	4.7487
Jason Bay	Pit	162	707	32	101	0.3055	0.4102	0.5593	4.5424
Adam Dunn	Cin	160	671	40	101	0.2468	0.3905	0.5396	4.3741
Vladimir Guerrero	LAA	141	594	32	108	0.3173	0.4057	0.5654	4.2820

[a]PA: plate appearance; HR: home runs; RBI: runs batted in; BA: batting average; OBP: on-base percentage; SLG: slugging percentage; WAS: winning advantage score.

strength and weakness. For each game, the difference in WAS a team accumulates (as a whole) between winning and losing is 1. Therefore, throughout a season, the total WAS a team accumulates is equal to the number of wins above 0.500. For example, in 2005 Houston Astros in the National League won 89 games and lost 73 games, so the accumulated WAS of all Astros players is equal to $89 - 81 = 8$. By breaking down this accumulated WAS into categories—pitching, batting, base running, and fielding—we can determine the strength and weakness of a team, as shown in Table 6. This analysis helps a general manager understand the needs in order to build a more balanced team.

Another usage of the winning advantage is to help a manager make on-field decisions. Specifically, in order to win, the manager should choose the play that maximizes the expected winning advantage. In baseball, examples include whether to attempt a steal, whether to lay a sacrifice bunt, and whether to intentionally walk a hitter. In basketball, examples include whether to attempt a 3-point shot, and whether to intentionally commit a foul.

4 Concluding Remarks

One may argue that using WAS to evaluate players is unfair to those players who do not get as many opportunities to perform when the game is on the line. We point out that the WAS does not measure a player's ability. Instead, it measures a player's contribution as a reflection from the game itself. A high WAS can be attributed to a player's luck, ability, or a combination of both—especially over a short period. However, over a long period such as an entire season, the luck tends to even out and the WAS does give a good indication of a player's ability.

Using WAS to evaluate players also has a distinctive benefit: It motivates the players to do what is best for the team. In sports, a player has an economic motivation to maximize his

Table 6: WAS breakdowns for National League teams in 2005.

Team	W	L	WAS by category				
			Pitching	Batting	Running	Fielding	Total[a]
SLN	100	62	13.1142	7.7369	1.0242	-2.8874	18.9879
ATL	90	72	2.1241	7.9306	1.1962	-2.2608	8.9900
HOU	89	73	15.8533	-6.0998	0.6037	-2.3571	8.0000
PHI	88	74	4.9704	2.5262	2.2931	-2.3787	7.4110
NYN	83	79	5.8568	-2.7128	2.2067	-3.3078	2.0428
FLO	83	79	8.3451	-3.6008	1.7814	-4.9267	1.5990
SDN	82	80	8.8969	-3.8681	0.2593	-4.2881	1.0000
MIL	81	81	3.0439	-0.2525	0.6297	-3.4210	0.0000
WAS	81	81	12.1820	-8.7111	-0.1590	-3.3546	-0.0428
CHN	79	83	2.8049	-0.5195	-0.1951	-4.0781	-1.9879
ARI	77	85	0.9314	-3.1203	0.9475	-2.7586	-4.0000
SFN	75	87	-0.4981	-3.1290	0.4060	-2.7788	-6.0000
CIN	73	89	-12.7067	6.1368	1.3123	-2.7424	-8.0000
LAN	71	91	-4.2703	-3.1512	0.4344	-3.0129	-10.0000
COL	67	95	-10.2653	-1.1624	0.8002	-3.3725	-14.0000
PIT	67	95	-3.3219	-8.2068	0.0942	-2.5655	-14.0000

[a]For some teams the total does not add up to an integer, because a few games were called before the bottom of the 9th inning due to inclement weather.

own market value by improving statistics that are appreciated in the open market. In many situations a certain "valuable" statistic does not contribute much to winning. If a player's reward (salary, bonus, or market value) is proportional to his WAS, a rational player should attempt to maximize the expected WAS in each play, which in turn maximizes the expected number of wins for his team.

Despite the usefulness of the winning advantage, it has two apparent drawbacks. First, the winning advantage is estimated from data, and the estimation—like all other estimations—is prone to errors. Even though a large set of historical play-by-play data is available, it is debatable whether the game is played the same way today as it was 50 years ago, so one cannot indefinitely improve the estimation by including more historical data. Second, for many plays it requires a subjective judgment to allocate WAS among responsible players. Unlike counting rebounds or steals in basketball, people may have different opinions about who contributes the most in a fast break. This disagreement makes the WAS a subjective statistic, and therefore undercuts its authority. In this paper, we make an effort to reduce these two drawbacks. We present an algorithm to improve the estimation of the winning advantage, and propose to divide WAS equally among responsible players when responsibility cannot be unambiguously determined. We believe the WAS calculated from these methods does provide useful insights into a player's contribution.

References

Bennett, J. 1993, Did shoeless Joe Jackson throw the 1919 world series?, *The American Statistician* **47**(4), 241–250.

Bennett, J. and Flueck, J. 1984, Player game percentage, *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 378–380.

Bennett, J. M. and Flueck, J. A. 1983, An evaluation of major league baseball offensive performance models, *The American Statistician* **37**(1), 76–82.

Berri, D. J. 1999, Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association, *Managerial and Decision Economics* **20**(8), 411–427.

Cover, T. M. and Keilers, C. W. 1977, An offensive earned-run average for baseball, *Operations Research* **25**(5), 729–740.

Hofler, R. A. and Payne, J. E. 1997, Measuring efficiency in the National Basketball Association, *Economics Letters* **55**(2), 293–299.

Mills, E. and Mills, H. 1970, *Player Win Averages*, A. S. Barnes, New York, NY.

Pankin, M. D. 1978, Evaluating offensive performance in baseball, *Operations Research* **26**(4), 610–619.

Ross, S. M. 2000, *Introduction to Probability and Statistics for Engineers and Scientists*, 2nd edn, Academic Press.

Zak, T. A., Huang, C. J. and Siegfried, J. J. 1979, Production efficiency: The case of professional basketball, *The Journal of Business* **52**(3), 379–392.