2006

# Effects of Home-Away Sequencing on the Length of Best-of-Seven Game Playoff Series

Christopher M. Rump
*Bowling Green State University*, cmrump@bgsu.edu

*Journal of Quantitative Analysis in Sports*

*Volume* 2, *Issue* 1 2006 *Article* 5

# The Effects of Home-Away Sequencing on the Length of Best-of-Seven Game Playoff Series

Christopher M. Rump*

*Bowling Green State University, cmrump@cba.bgsu.edu

# The Effects of Home-Away Sequencing on the Length of Best-of-Seven Game Playoff Series

Christopher M. Rump


## Abstract

We analyze the number of games played in a seven-game playoff series under various home-away sequences. In doing so, we employ a simple Bernoulli model of home-field advantage in which the outcome of each game in the series depends only on whether it is played at home or away with respect to a designated home team. Considering all such sequences that begin and end at home, we show that, in terms of the number of games played, there are four classes of stochastically different formats, including the popular 2-3 and 2-2 formats both currently used in National Basketball Association (NBA) playoffs. Characterizing the regions in parametric space that give rise to distinct stochastic and expected value orderings of series length among these four format classes, we then investigate where in this parametric space that teams actually play. An extensive analysis of historical 7-game playoff series data from the NBA reveals that this home-away model is preferable to the simpler, well-studied but ill-fitting binomial model that ignores home-field advantage. The model suggests that switching from the 2-2 series format used for most of the playoffs to the 2-3 format that has been used in the NBA Finals since a switch in 1985 would stochastically lengthen these playoff series, creating an expectation of approximately one extra game per playoff season. Such evidence should encourage television sponsors to lobby for a change of playoff format in order to garner additional television advertising revenues while reducing team and media travel costs.


**KEYWORDS:** Playoffs, Tournament, Sequencing, Probability

# 1 Introduction

This article studies a model of multi-game playoff series pairing two sports teams, one of whom is favored with a home-field advantage. The advantage means that a majority of the possible games in the series are played on that team's home field. We focus on the case of seven-game series, which have been long adopted by the major professional team sports championships. Here four games are scheduled to be played on the favored team's home field, the other three games to be played away from this home field. Of course all seven games need not always be played since these series are played until one team wins a majority of the games, i.e., four games. The issue of home-away sequencing then naturally arises. If playing at home actually gives an advantage, then this sequencing will play a role in determining the number of games actually played. For example, if the home team enjoys a large probability of winning, then a series beginning with four straight games at home would give a tremendous (and deemed unfair) advantage to that team to win a short series.

Beginning with the seminal work of Mosteller (1952), many authors, including Brunner (1987), Groeneveld and Meeden (1975), Nahin (2000), Simon (1977), Woodside (1989), and Zelinski (1973), have analyzed a simple truncated binomial model of the World Series or other best-of-seven game playoff series, in which one team is assigned a probability $p$ of winning any particular game in the series. This model assumes that each game is played independently of the next, and without regard for widely assumed home-team advantage. Lengyel (1993) extended the analysis of this binomial model to find an expression for the expected length of any best-of-$(2n-1)$ game series, although a more elegant form was already presented by Maisel (1966).[1] Shapiro and Hamilton (1993) rediscovered Maisel's elegant result and further recognized that this expression surprisingly involves partial sums of the Catalan numbers, $C_k = \binom{2k}{k}/(k+1)$, $k = 1, 2, \ldots$.

Such best-of-seven playoff tournaments, originally conceived to determine the best of two teams from different leagues, are one form of paired statistical comparison as studied by David (1988), Uppuluri and Blot (1974) and Ushakov (1976). Of course, as explained in James et al. (1993), the best teams do not always come out ahead in such comparisons. As a result, many authors, including Appleton (1995), Boronico (1999), Gibbons et al. (1978), Glenn (1960), McGarry (1998) and most recently Marchand (2002), have studied the

[1] Mosteller (1952) earlier derived the same elegant form for the special case of the World Series when adding the $n = 4$ wins of the series winner to the expected number of wins by the series loser as expressed in his equation (2).

effectiveness of a variety of multi-tiered playoff tournaments among several playoff teams.

Although the existence of a home-team advantage in sporting events has been documented in recent years by Clarke and Norman (1995), Courneya and Carron (1992) and Harville and Smith (1994), little modelling of this effect has taken place, perhaps tempered by the early analysis of Mosteller (1952) that revealed no significant home advantage in the World Series, at least through the first half-century of play. Bassett and Hurley (1998) appear to be the first to extend the common single-parameter binomial model to a two-parameter model incorporating home and away winning probabilities $p_H$ and $p_A$, respectively, for the team with home-field advantage. One of their chief results involves the relatively simple conditions under which the so-called 2-3 playoff format (HHAAAHH) is longer (in expectation) than the 2-2 format (HHAAHAH).

Our work extends this analysis by Bassett and Hurley to include all possible home-away sequences involving 7 games, four of which are at played at home including the first and seventh game of the series. We first review the home-away model in Section 2. Then, in Section 3 we show that, in terms of the number of games played, there are four stochastically different format classes, including the popular 2-3 and 2-2 formats. We then compare these four formats in terms of both stochastic and expected length. The expected value orderings among these four format classes partition the two-dimensional parametric space into six regions.

The remainder of the paper then validates the effectiveness of the model in explaining the outcomes of playoffs in the National Basketball Association (NBA). In Section 4, we examine the historical data to first estimate the region in which the parameters actually lie, and then study the goodness of fit of the simple binomial model versus the home-away model. We close in Section 5 with a discussion of our results and a prelude of a more general Markov model of game-to-game dependence in playoff series.

## 2 Model

Following Bassett and Hurley (1998), suppose each game in a sequence of playoff games is an independent Bernoulli trial with probability of success $p_H$ if the game is played at home and probability of success $p_A$ if the game is played away from home. Here “home” refers to the apparently favored team’s court/field/ice and “success” refers to a win by this favored team. The favored team is designated as the team with the home advantage, meaning that four of the possible seven games in the series are played on that team's home court. Since the home court usually gives an advantage, $p_H \geq p_A$ is typically the case.

We will restrict such series to the convention that the series should begin and end at home, i.e., the first and seventh (if necessary) games are scheduled on the home team's court. With this restriction it is possible to sequence the middle five games in $\binom{5}{2} = 10$ possible ways as listed in Table 1.

| Format | Sequence |
|---|---|
| 2-3 | HHAAAHH |
| 2-2 | HHAAHAH |
| 2-1 | HHAHAAH |
| 1-1-2 | HAHHAAH |
| 3-3 | HHHAAAH |

| Format | Sequence |
|---|---|
| 1-1 | HAHAHAH |
| 1-2 | HAHAAHH |
| 1-2-1 | HAAHAHH |
| 1-2-2 | HAAHHAH |
| 1-3 | HAAAHHH |

Table 1: Possible Home-Away Sequences in a 7-Game Series

The first entry in each of the two lists is a palindromic sequence. Each of the remaining sequences in the right-hand list is simply the reverse of the corresponding sequence in the left-hand list. The format indicates how the sequence alternates at the beginning of the series, from which the entire sequence can be inferred. For example, the format 2-1 indicates two home games followed by one away game. From this we infer that the fourth game is home again. Since the last game must be home as well, this implies that games five and six are away. For brevity the formats 1-1-1-1 and 1-1-1-2 have been shortened to 1-1 and 1-2, respectively.

In terms of practical use, the National Hockey League (NHL) playoffs primarily employ the 2-2 format, although, according to National Hockey League (1997), under certain geographic conditions the higher-ranked team has the option to invoke a 2-3 format in order to limit travel. Since 1924, Major League Baseball's (MLB) World Series has used the 2-3 format. In the first 20 or so years of that series, organizers tinkered with many seven-game formats such as 1-1, 2-2 and 1-2-2. In the National Basketball Association (NBA) Finals, the 2-3 format has been used since 1985. Prior to that, the 2-2 format was predominant, although the 1-1 and 1-2-2 formats were also used from time to time. The rest of the NBA playoffs (preceding the NBA Finals) are played in the 2-2 format, although in the early days the eastern conference playoffs often used a 1-1 format.

# 3 Series Length

We shall now explore the effects of the home-away sequencing on the length of a series. Following the notation of Bassett and Hurley (1998), for each format $f$ let $h_i^f$ ($a_i^f$) be the probability that the home (away) team wins the series in $i$ games, and $g_i^f = h_i^f + a_i^f$ be the probability that the series is won in $i$ games, $i = 4, 5, 6, 7$. Also, let $h^f = h_4^f + h_5^f + h_6^f + h_7^f$ be the probability that the home team wins the series under format $f$, and let $a^f = 1 - h^f$ be the probability that the away team does so.

Bassett and Hurley argue that, due to the independence of games, the probability that the home team wins the series is the same for any format, i.e, the probabilities $h^f$ (and $a^f$) are invariant to the format $f$. Extending the argument, they state that the probability that the series lasts seven games is the same for both the 2-3 and 2-2 formats, i.e., $g_7^{23} = g_7^{22}$. This is also due to the independence of games and the fact that both formats end with a seventh game at home, i.e., the first six games contain the same mix of home and away games. By independence, the particular mix does not affect the probability that neither team wins four of these six games, warranting a final seventh game. Since we assume that the seventh game of a series is always played at home, this result extends to all the formats in Table 1.

To explore the length of the playoff series under different formats, let $N_f$ indicate the number of games played under the format $f$, and let $L_f = E[N_f]$ be its expectation. Since each playoff series consists of at least four games, and the games are assumed to be independent, the sequencing of the last three games will affect the number of games played. Hence, those sequences with the same sequence in the last three games, and therefore the same mix of home and away games over the first four requisite games, will be stochastically equivalent, denoted $=_{\text{st}}$. [2] This establishes Theorem 1.

**Theorem 1** *Assuming the games in a best 4 of 7-game series are independent, only the sequencing of the last three games will affect the number of games played. Thus, under this assumption, series with the same home-away sequence over the last 3 games are stochastically equivalent in length. Hence,*

- $N_{23} =_{\text{st}} N_{12} =_{\text{st}} N_{121}$,
- $N_{22} =_{\text{st}} N_{11} =_{\text{st}} N_{122}$,
- $N_{21} =_{\text{st}} N_{33} =_{\text{st}} N_{112}$.

[2]For a discussion of stochastic order relations, see Ross (1996).

In light of Theorem 1, we will focus on the four stochastically different formats 2-3, 2-2, 2-1 and 1-3. We have already established that, since the last game of each of these formats ends in a home game, the series have the same probability of lasting seven games, i.e.,

$$g_7^{23} = g_7^{22} = g_7^{21} = g_7^{13}. \tag{1}$$

Since the last two games of formats 2-3 and 1-3 are both home games, $g_6^{23} + g_7^{23} = g_6^{13} + g_7^{13}$, yielding via (1)

$$g_6^{23} = g_6^{13}. \tag{2}$$

Similarly, the last two games of formats 2-2 and 2-1 are identical so that

$$g_6^{22} = g_6^{21}. \tag{3}$$

Extending this analysis to include the last three games gave us Theorem 1. We also have

$$g_4^{23} = g_4^{22}, \tag{4}$$

as these two formats both have identical structure over the first four games. As a consequence of Equations (1)-(4), we have the following identities:

$$\begin{aligned} g_4^{23} + g_5^{23} &= g_4^{13} + g_5^{13} && (5)\\ g_4^{22} + g_5^{22} &= g_4^{21} + g_5^{21} && (6)\\ g_5^{23} + g_6^{23} &= g_5^{22} + g_6^{22} && (7) \end{aligned}$$

Bassett and Hurley show that

$$L_{23} - L_{22} = 2(p_H - p_A)(p_H + p_A - 1)(p_H + p_A - 2p_H p_A).$$

Hence, since the last term $p_H + p_A - 2p_H p_A$ is bounded between 0 and 1, a 2-3 series is longer than a 2-2 series in expectation, $L_{23} \geq L_{22}$, if and only if $p_H - p_A$ agrees in sign with

$$\varphi(p_H, p_A) = \varphi(p_A, p_H) = p_H + p_A - 1.$$

Assuming $p_H \geq p_A$, $L_{23} \geq L_{22}$ holds provided that $1 - p_H \leq p_A \leq p_H$. It turns out that Bassett and Hurley have proven a stronger result, namely that these are the conditions for a 2-3 series to be stochastically longer than a 2-2 series, $N_{23} \geq_{\text{st}} N_{22}$. This follows from their derivation

$$\begin{aligned} L_{23} - L_{22} &= 5(g_5^{23} - g_5^{22}) + 6(g_6^{23} - g_6^{22}) \\ &= 5((g_5^{23} + g_6^{23}) - (g_5^{22} + g_6^{22})) + (g_6^{23} - g_6^{22}) \\ &= g_6^{23} - g_6^{22} = g_5^{22} - g_5^{23}, \end{aligned}$$

where the last two equalities follow from (7).

In words, the 2-3 format delays the third home game until game six rather than game five. Thus, assuming $p_H \geq p_A$, the home team is less likely to win game five under the 2-3 format. This will delay the home team from winning the series - and thereby lengthen the series - provided that this does not allow the other team to win the series earlier, which can happen if the home team is relatively weak on the road, i.e., $p_A < 1 - p_H$.

Comparing the 2-2 and 2-1 formats, we see that the difference lies in games 4 and 5. Thus, a 2-2 series is stochastically longer than a 2-1 series, $N_{22} \geq_{\rm st} N_{21}$, if and only if

$$\begin{aligned} 0 \leq L_{22} - L_{21} &= g_5^{22} - g_5^{21} = g_4^{21} - g_4^{22} = (h_4^{21} - h_4^{22}) + (a_4^{21} - a_4^{22}) \\ &= (p_H - p_A)p_H^2 p_A + [(1-p_H) - (1-p_A)](1-p_H)^2(1-p_A) \\ &= (p_H - p_A)\left[p_H^2 p_A - (1-p_H)^2(1-p_A)\right]. \end{aligned}$$

Thus, $N_{22} \geq_{\rm st} N_{21}$ (and $L_{22} \geq L_{21}$) if and only if $p_H - p_A$ agrees in sign with

$$\phi(p_H, p_A) = p_H^2 p_A - (1-p_H)^2(1-p_A).$$

Assuming $p_H \geq p_A$, $N_{22} \geq_{\rm st} N_{21}$ holds provided $\dfrac{(1-p_H)^2}{p_H^2 + (1-p_H)^2} \leq p_A \leq p_H$.

Comparing the 1-3 and 2-3 formats would seem to be a bit trickier. However, due to Theorem 1, we can compare the 1-3 format to the 1-2-1 format instead. These formats, like in the comparison of the 2-2 and 2-1 formats, differ only in games 4 and 5. However, unlike the 2-2 and 2-1 formats, the 1-3 and 2-3 formats contain two away games rather than home games in their first three games. Hence, $N_{13} \geq_{\rm st} N_{23}$ (and $L_{13} \geq L_{23}$), if and only if $\phi(p_A, p_H)$ and $p_H - p_A$ agree in sign. Here $\phi(p_A, p_H)$ is a reflection of $\phi(p_H, p_A)$ across the line $p_A = p_H$.

The previous pairwise comparisons each involved formats that differed in adjacent games. In comparing the length of series that differ in non-adjacent games, we first combine the previous comparisons to make statements about dominance in expected value. For example, comparing formats 2-3 and 2-1, we find

$$\begin{aligned} L_{23} - L_{21} &= (L_{23} - L_{22}) + (L_{22} - L_{21}) \\ &= (p_H - p_A)\left[2(1-2p_H)p_A^2 - (1 - 6p_H + 2p_H^2)p_A - (1-p_H^2)\right]. \end{aligned}$$

Thus, $L_{23} \geq L_{21}$ holds if and only if $p_H - p_A$ agrees in sign with

$$\psi(p_H, p_A) = 2(1-2p_H)p_A^2 - (1 - 6p_H + 2p_H^2)p_A - (1-p_H)(1+p_H).$$

Similarly, $L_{13} \geq L_{22}$ holds if and only if $\psi(p_A, p_H)$ and $p_H - p_A$ agree in sign since $L_{13} - L_{22} = (L_{23} - L_{22}) + (L_{13} - L_{23})$ is a reflection of $L_{23} - L_{21}$ across the line $p_A = p_H$. Finally,

$$\begin{aligned} L_{13} - L_{21} &= (L_{13} - L_{22}) + (L_{22} - L_{21}) = (p_H - p_A)[\psi(p_A, p_H) - \phi(p_H, p_A)] \\ &= (L_{13} - L_{23}) + (L_{23} - L_{21}) = (p_H - p_A)[\phi(p_A, p_H) - \psi(p_H, p_A)] \\ &= (p_H - p_A)(p_H + p_A - 1)[p_H + p_A + 2(1 - p_H p_A)], \end{aligned}$$

which is non-negative if and only if $p_H - p_A$ and $p_H + p_A - 1$ agree in sign. Hence, a 1-3 series is longer in expectation than a 2-1 series under the same conditions for which a 2-3 series is longer in expectation than a 2-2 series.

Next, we consider stochastic ordering of the formats that differ in non-adjacent games. Comparing the 2-3 and 2-1 formats, $N_{23} \geq_{\text{st}} N_{21}$ holds if and only if

$$g_4^{21} \geq g_4^{23}, \tag{8}$$

$$g_4^{21} + g_5^{21} \geq g_4^{23} + g_5^{23}, \tag{9}$$

$$g_4^{21} + g_5^{21} + g_6^{21} \geq g_4^{23} + g_5^{23} + g_6^{23}. \tag{10}$$

Condition (8) becomes $g_4^{21} \geq g_4^{22}$ using (4). This is equivalent to $N_{22} \geq_{\text{st}} N_{21}$. Condition (9) becomes $g_5^{22} \geq g_5^{23}$ using (6) and (4). This is equivalent to $N_{23} \geq_{\text{st}} N_{22}$. Condition (10) holds from (1). Hence, $N_{23} \geq_{\text{st}} N_{21}$ holds if and only if $N_{23} \geq_{\text{st}} N_{22} \geq_{\text{st}} N_{21}$. In similar fashion it is possible to show that $N_{13} \geq_{\text{st}} N_{22}$ holds if and only if $N_{13} \geq_{\text{st}} N_{23} \geq_{\text{st}} N_{22}$.

Finally, comparing the 1-3 and 2-1 formats, $N_{13} \geq_{\text{st}} N_{21}$ holds if and only if

$$g_4^{21} \geq g_4^{13}, \tag{11}$$

$$g_4^{21} + g_5^{21} \geq g_4^{13} + g_5^{13}. \tag{12}$$

Using (6), (5) and (4), Condition (12) becomes $g_5^{22} \geq g_5^{23}$ or $N_{23} \geq_{\text{st}} N_{22}$. Condition (11) holds if and only if

$$(p_H - p_A)(p_H + p_A - 1)(2 - p_H - p_A + 2p_H p_A) \geq 0$$

or, since the last term is always positive, simply $N_{23} \geq_{\text{st}} N_{22}$. Hence, $N_{13} \geq_{\text{st}} N_{21}$ holds if and only if $N_{23} \geq_{\text{st}} N_{22}$.

The regions of stochastic and expected value ordering of these playoff formats are summarized in Table 2 and illustrated in Figure 1. Theorem 2 summarizes these results.

| Parameter Region | Agrees in sign with $p_H - p_A$ | | | | |
|---|---|---|---|---|---|
| | $\phi(p_A, p_H)$ | $\psi(p_A, p_H)$ | $\varphi(p_H, p_A)$ | $\psi(p_H, p_A)$ | $\phi(p_H, p_A)$ |
| A | $\surd$ | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| B | | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| C | | | $\surd$ | $\surd$ | $\surd$ |
| c | | | | $\surd$ | $\surd$ |
| b | | | | | $\surd$ |
| a | | | | | |

Table 2: Parametric Region Definitions

**Theorem 2** *In a best 4 of 7-game series with independent games, the home and away sequencing formats influence the number of games played. The lengths of a series under various formats are ordered stochastically and in expectation according to Table 3.*

| Parameter Region | Stochastic Ordering | Expected Value Ordering |
|---|---|---|
| A | $N_{13} \geq_{\text{st}} N_{23} \geq_{\text{st}} N_{22} \geq_{\text{st}} N_{21}$ | $L_{13} \geq L_{23} \geq L_{22} \geq L_{21}$ |
| B | $N_{23} \geq_{\text{st}} N_{13} \geq_{\text{st}} N_{21}$<br>$N_{23} \geq_{\text{st}} N_{22} \geq_{\text{st}} N_{21}$ | $L_{23} \geq L_{13} \geq L_{22} \geq L_{21}$ |
| C | $N_{23} \geq_{\text{st}} N_{13} \geq_{\text{st}} N_{21}$<br>$N_{23} \geq_{\text{st}} N_{22} \geq_{\text{st}} N_{21}$ | $L_{23} \geq L_{22} \geq L_{13} \geq L_{21}$ |
| c | $N_{22} \geq_{\text{st}} N_{23} \geq_{\text{st}} N_{13}$<br>$N_{22} \geq_{\text{st}} N_{21} \geq_{\text{st}} N_{13}$ | $L_{22} \geq L_{23} \geq L_{21} \geq L_{13}$ |
| b | $N_{22} \geq_{\text{st}} N_{23} \geq_{\text{st}} N_{13}$<br>$N_{22} \geq_{\text{st}} N_{21} \geq_{\text{st}} N_{13}$ | $L_{22} \geq L_{21} \geq L_{23} \geq L_{13}$ |
| a | $N_{21} \geq_{\text{st}} N_{22} \geq_{\text{st}} N_{23} \geq_{\text{st}} N_{13}$ | $L_{21} \geq L_{22} \geq L_{23} \geq L_{13}$ |

Table 3: Playoff Series Length Orderings by Parametric Region

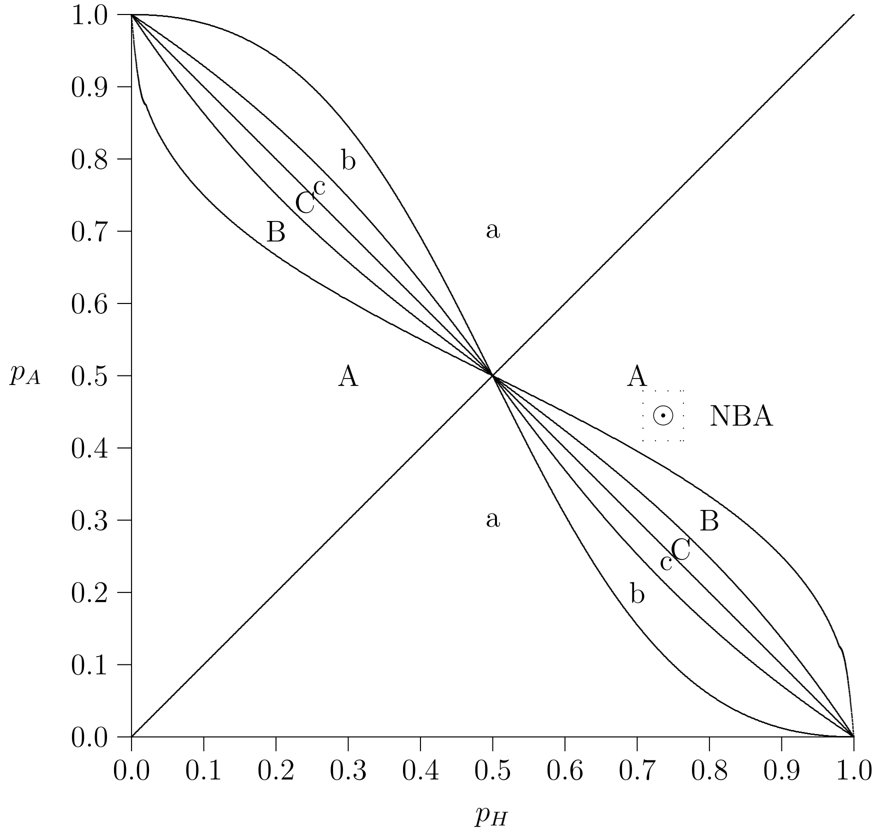


Figure 1: Parametric Region Partition. The Curves Separate Regions - Defined in Table 2 - that Characterize Different Series-Length Orderings for the Various Home-Away Format Families.

The series formats do not lend themselves to stronger orderings as described in Ross (1996). For example, the lengths of the series formats are not likelihood ratio ordered as a consequence of (1). Equations (5)-(7) reveal that the lengths of the series are not hazard rate ordered either. Here we define the hazard rate of the discrete random variable $N_f$ at game $i$ to be $g_i^f \Big/ \left(1 - \sum_{j=4}^{i-1} g_j^f\right)$, $i = 4, 5, 6, 7$. We interpret this "rate" to be the probability that the series will end in game $i$ given that a series has reached game $i$.

# 4 Model Validation

To validate the home-away sequencing model, we collected game-by-game playoff data from National Basketball Association (NBA) historical records in Hubbard (2000). We supplemented these records with the results of more recent play (through year 2005) as documented at the league web site www.nba.com.

The NBA data includes 305 playoff series that were played in the 2-2 format family (i.e., the 2-2, 1-1 or 1-1-2 formats). These series are comprised of first-round, conference semi-final, conference final and NBA final playoffs. Except for the years 1949 and 1953-55 of 2-3 format play, the NBA Finals used the 2-2 format family from their inception in 1947 through 1984, after which a 2-3 format has been adopted. This represents 34 playoff series under the 2-2 format family. The NBA conference finals (in both the Eastern and Western conferences) have been played in this format family since their inception in 1958, which provides $2 \times 48 = 96$ additional playoff series. Similarly, all but one of the conference semi-finals have been played in this format family since their inception in 1968, which provides $4 \times 38 - 1 = 151$ additional playoff series.[3] Very recently (starting in 2003), the eight first-round series of the NBA playoffs were extended to 7-game series. This provides $3 \times 8 = 24$ additional playoff series, for a grand total of 305. [4]

There have been 1744 games played in these 305 series, for an average series length of 5.718 games. This is just under the maximum expected length of 5.8125 games between two evenly matched teams under a binomial model as shown by Brunner (1987), Groeneveld and Meeden (1975), Nahin (2000) and Woodside (1989). Of these 1744 games, the favored team won 1055, for a winning percentage of about 60.5%. Thus, a straight-forward estimate of the Bernoulli success probability $p$ is $\hat{p} = 1055/1744 \approx 0.605$. [5]

For the home-away model, we compute similar estimates based on whether the favored team was at home or away. These estimates are $\hat{p}_H = 706/959 \approx 0.736$ and $\hat{p}_A = 349/785 \approx 0.445$. A simple 95% confidence interval about

[3] The 1971 western semi-finals series between San Francisco and Minneapolis was played under the 1-2-1 format.

[4] It is too early to tell if this limited number of first-round series are significantly different than the later playoff series that typically pair teams of more equal caliber. We include them for completeness and note that their exclusion changes the results only slightly.

[5] Mosteller (1952, 1997) discusses several estimation procedures for the probability that the "better" team wins each game, where the "better" team is not known. Since we need estimates of the strength of the "favored" team, which is known, we do not resort to these more complicated methods. Further, the primary methods of Mosteller (1952), namely, maximum likelihood estimation and minimum chi-square estimation, provide nearly identical results.

each of these independent Bernoulli parameter estimates, $\hat{p}_H$ and $\hat{p}_A$, produces an approximate 90% confidence region about the joint point estimate $(\hat{p}_H, \hat{p}_A)$, which is completely contained within parameter region A (where the 1-3 format dominates), as shown in Figure 1. [6] The location of this confidence region suggests that a simple binomial model is not appropriate since this region does not cross the line $\hat{p}_A = \hat{p}_H$. Further, it suggest that, although played under the 2-2 family format, these playoff series would have been theoretically longer if they had been played under a 1-3 format or, as are the recent NBA championships, a 2-3 format.

Table 4 presents the Pearson ($\chi^2$) goodness of fit of the various models to the NBA playoff data outcomes (cf., e.g., Sec. 14.2 of Devore (2004)). The first eight columns of numbers indicate the number of the 305 series that would result in the series outcomes $x$-$y$, where $x$ and $y$ represent the number of games won by the home and away team, respectively. The first row of the table indicates the actual observed frequencies of such outcomes and the remaining rows list the expected number of such outcomes as predicted by the various models.

| | Outcome | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model | 4-0 | 4-1 | 4-2 | 4-3 | 3-4 | 2-4 | 1-4 | 0-4 | p |
| Actual | 33 | 74 | 53 | 71 | 15 | 36 | 14 | 9 | 1.000 |
| Binomial | 40.8 | 64.5 | 63.7 | 50.4 | 32.9 | 27.2 | 17.0 | 7.4 | 0.000 |
| 1-3 Family | 19.7 | 59.6 | 80.2 | 66.7 | 23.9 | 22.2 | 18.9 | 13.8 | 0.000 |
| 2-3 Family | 32.7 | 46.7 | 80.2 | 66.7 | 23.9 | 22.2 | 26.1 | 6.5 | 0.000 |
| 2-2 Family | 32.7 | 77.3 | 49.5 | 66.7 | 23.9 | 35.9 | 12.4 | 6.5 | 0.403 |
| 2-1 Family | 54.1 | 55.9 | 49.5 | 66.7 | 23.9 | 35.9 | 15.8 | 3.1 | 0.000 |

Table 4: Goodness of Fit of Models in Predicting the Outcomes of 305 NBA 7-Game Playoff Series Played in a 2-2 Home-Away Format.

The last column of Table 4 reports the significance probability (the so-called p-value) for the chi-squared goodness-of-fit test of the models. For each model, this p-value represents the probability that we would observe outcome numbers that deviate at least as much as the actual observed figures given

[6]The horizontal confidence interval was computed as $\hat{p}_H \pm z \cdot \hat{\sigma}$ with $z = 1.96$ and standard deviation estimate $\hat{\sigma} = \sqrt{\hat{p}_H(1-\hat{p}_H)/n_H}$, where $n_H = 959$ is the total number of home games played. A similar vertical confidence interval was computed for the away games.

that the model perfectly characterizes the randomness of the series outcome. Therefore, small p-values near 0 indicate a poor model (that we reject as a good fit for the data), whereas large p-values near 1 indicate a good model. [7]

Examining Table 4 we see that, although parsimonious, the binomial model does not fit the data. On the other hand, the 2-2 home-away model fits the data fairly well with a p-value over 0.4. As should be expected, the other home-away models do not fit the data since the series were not played under those formats.

There is a fairly large overestimation of 3-4 outcomes by all the models, partly due to the fact that of the 86 series that have gone to a full 7 games, the favored team (at home) has won 71 of these 86 final games for a winning probability of 0.826, which is somewhat larger than the home-away models' probability estimate $\hat{p}_H = 0.736$.

Satisfied with the fit of the home-away model to the actual outcomes $x$-$y$, we can now examine effects on the series length $x+y$. Table 5 presents the observed and expected relative frequencies for the NBA playoff series length. The last column lists the average number of games played in the series, computed by summing the products of each length by its relative frequency.

| | Series Length | | | | |
|---|---|---|---|---|---|
| Model | 4 | 5 | 6 | 7 | Average |
| Actual | 0.138 | 0.289 | 0.292 | 0.282 | 5.718 |
| Binomial | 0.158 | 0.271 | 0.298 | 0.273 | 5.686 |
| 1-3 Family | 0.110 | 0.257 | 0.336 | 0.297 | 5.820 |
| 2-3 Family | 0.129 | 0.239 | 0.336 | 0.297 | 5.801 |
| 2-2 Family | 0.129 | 0.294 | 0.280 | 0.297 | 5.746 |
| 2-1 Family | 0.188 | 0.235 | 0.280 | 0.297 | 5.687 |

Table 5: Relative Frequencies for the Length of NBA 7-Game Series.

The actual frequencies of series length work out to roughly 1 out of every 7 series ending in a 4-game "sweep", and a uniform 2 out of 7 chance for each of the other three possible series lengths. Thus, in every postseason, which consists of 14 playoff series leading up to the NBA Finals, roughly 2 of the

[7] The number of degrees of freedom used for the chi-squared goodness-of-fit test is the number of categories less the number of parameters estimated (plus 1). For the binomial model with parameter estimate $p$, the degrees of freedom are 8-(1+1)=6. Since the home-away models require two parameters, $p_H$ and $p_A$, the degrees of freedom are 5.

series are sweeps, 4 series are of length 5 games, 4 are of length 6, and 4 are full series of length 7. The 2-2 model does the best job at replicating these frequencies and is closest in matching the average length.

As indicated by Table 5, if these playoffs were played under the 2-3 format - which since 1985 has been used exclusively for the NBA finals - the expected series length would increase from 5.746 games under the 2-2 format to 5.801 games under a 2-3 format, an increase of 0.055 games per series. Since these formats differ only in a swapping of the locations of games 5 and 6, this difference of 0.055 is precisely the probability mass that is shifted from a 5-game series outcome under a 2-2 format to a 6-game outcome instead under the 2-3 format. [8]

This increase amounts to an extra game in one of every $1/0.055 \approx 18$ series. Since 14 such playoff series are played every year now, one could expect almost one extra playoff game added to the current average of 80 or so ($\approx 14 \cdot 5.746$) games played each year leading up to the NBA finals. This is fairly small effect, but might be sufficiently lucrative to the television networks who pay huge fees for the broadcast rights to these games.

As can be discerned from Table 4, the expected result of this switch in formats would be a significant (over 60%) increase in 4-2 outcomes, most of which were formerly 4-1 series wins by the favored team at home in game 5 under the 2-2 format. This predicted increase is offset somewhat by a corresponding increase in shorter 1-4 outcomes, where the away team wins the series with the advantage at home in game 5 rather than in game 6 of a 2-4 outcome.

The 2-3 format would also seem attractive in cutting down on travel "costs," both financial (impacting team and media budgets) and physical (impacting team and media fatigue). The 2-2 format requires up to ten trips, four by the favored home team and six by the away team. In contrast, the 2-3 format requires at most six trips, only two by the favored home team and four by the away team.

A 1-3 format would, in theory, extend the number of games even more than the 2-3 format, with no additional travel cost. However, the perceived psychological advantage to the away team (with a guarantee to host all three of their games) and lack of precedent would likely preclude adoption of such a format.

[8]The 25 NBA Finals played under the 2-3 format have averaged 5.720 games. Thus, this limited data has yet to reveal any significant difference in playoff series length from the 5.718 games listed in Table 5 for the 2-2 format family.

# 5 Conclusions

We have presented an extended analysis of the number of games played in a playoff series under various home-away sequences. Using the model of Bassett and Hurley (1998), we considered all possible playoff sequences in which the series begins and ends on the home team's court. Under this model it turns out that, in terms of the number of games played, these sequences reduce to four stochastically different formats, including the popular 2-3 and 2-2 formats both currently used in National Basketball Association (NBA) playoffs. We then characterized the parametric regions that give rise to six distinct stochastic and expected value orderings among these four format classes.

Examining the performance of the home-away model using National Basketball Association (NBA) playoff data, we found that the home-away model fits the data fairly well, whereas the simple binomial model so prevalent in the literature did not. We also discovered that the NBA could stochastically extend the length of the conference championships by switching from the long-standing 2-2 series format to a 2-3 format, as was done for the NBA finals in 1985. Such evidence should encourage television sponsors to lobby the NBA for a change of playoff format in order to garner additional television advertising revenues while reducing team and media travel costs.

In contrast, similar analysis of the National Hockey League (NHL) 7-game playoff data reveals that neither the binomial model nor the home-away model provide any explanatory power. In these playoffs there is a prevalence of many short "sweeped" series where the game outcome seems heavily influenced by the presence of a team facing elimination. This violates the assumption of game independence assumed throughout this paper. However, a two-parameter Markov chain formulation, built around the home-away model performs extremely well as reported in Rump (2005).

# References

D. R. Appleton. May the best man win? *The Statistician*, 44(4):529–538, 1995.

G. W. Bassett and W. J. Hurley. The effects of alternative HOME-AWAY sequences in a best-of-seven playoff series. *The American Statistician*, 52 (1):51–53, 1998.

J. S. Boronico. Multi-tiered playoffs and their impact on professional baseball. *The American Statistician*, 53(1):56–61, 1999.

J. Brunner. Absorbing Markov chains and the number of games in a World Series. *The UMAP Journal*, 8(2):99–108, 1987.

S. R. Clarke and J. M. Norman. Home ground advantage of individual clubs in English soccer. *The Statistician*, 44(4):509–521, 1995.

K. S. Courneya and A. V. Carron. The home advantage in sport comptetitions: a literature review. *Journal of Sport Exercise Psychology*, 14:13–27, 1992.

H. A. David. *The Method of Paired Comparisons*. Charles Griffin & Company, London, 1988.

J. L. Devore. *Probability and Statistics for Engineering and the Sciences*. Thomson Brooks/Cole, 6 edition, 2004.

J. D. Gibbons, I. Olkin, and M. Sobel. Baseball competitions - are enough games played? *The American Statistician*, 32(3):89–95, 1978.

W. A. Glenn. A comparison of the effectiveness of tournaments. *Biometrika*, 47(3-4):253–262, 1960.

R. A. Groeneveld and G. Meeden. Seven game series in sports. *Mathematics Magazine*, 48(4):187–192, 1975.

D. A. Harville and M. H. Smith. The home-court advantage: How large is it, and does it vary from team to team? *The American Statistician*, 48(1): 22–28, 1994.

J. Hubbard, editor. *The Official NBA Encyclopedia*. Doubleday, New York, 3rd edition, 2000.

B. James, J. Albert, and H. S. Stern. Answering questions about baseball using statistics. *Chance*, 6(2):17–22, 30, 1993.

T. Lengyel. A combinatorial identity and the World Series. *SIAM Review*, 35 (2):294–297, 1993.

H. Maisel. Best $k$ of $2k-1$ comparisons. *Journal of the American Statistical Association*, 61:329–344, 1966.

É. Marchand. On the comparison between standard and random knockout tournaments. *Statistician*, 51(2):169–178, 2002.

T. McGarry. On the design of sports tournaments. In J. Bennett, editor, *Statistics in Sport*, Arnold Applications of Statistics, chapter 10, pages 199–217. Arnold, London, 1998.

F. Mosteller. The World Series competition. *Journal of The American Statistical Association*, 47(259):355–380, 1952.

F. Mosteller. Lessons from sports statistics. *The American Statistician*, 51(4): 305–310, 1997.

P. J. Nahin. *Duelling Idiots and other Probability Puzzlers.* Princeton University Press, Princeton, NJ, 2000.

National Hockey League. *Stanley Cup Playoffs Fact Guide 1998.* Triumph Books, Chicago, 1997.

S. M. Ross. *Stochastic Processes.* John Wiley & Sons, New York, 2nd edition, 1996.

C. M. Rump. Andrei Markov in the Stanley Cup playoffs. *Chance Magazine*, 2005. In press.

L. Shapiro and W. Hamilton. The Catalan numbers visit the World Series. *Mathematics Magazine*, 66(1):20–22, 1993.

W. Simon. Back-to-the-wall effect: 1976 perspective. In S. P. Ladany and R. E. Machol, editors, *Optimal Strategies in Sports*, volume 5 of *Studies in Management Science and Systems*, Amsterdam, 1977. North-Holland.

V. R. R. Uppuluri and W. J. Blot. Asymptotic properties of the number of replications of a paired comparison. *Journal of Applied Probability*, 11: 43–52, 1974.

I. A. Ushakov. The problem of choosing the preferred element: An application to sport games. In R. E. Machol, S. P. Ladany, and D. G. Morrison, editors, *Management Science in Sports*, volume 4 of *Studies in the Management Sciences*, Amsterdam, 1976. North-Holland.

W. Woodside. Winning streaks, shutouts, and the length of the World Series. *The UMAP Journal*, 10(2):99–113, 1989.

M. Zelinski. How many games to complete a World Series. In F. Mosteller, editor, *Weighing Chances*, Statistics by Example: Weighing Chances, Reading, 1973. Addison-Wesley.