**Studying Baseball’s Wild-Card Team Using Probability**

**Richard A. Auer and Michael P. Knapp**

**1. Introduction**

 The modern era of professional baseball playoffs began in 1903 when the champions of the American League and the National League played in the first World Series. Except for 1904, this playoff system was maintained until 1969. From 1969 through 1993, each of the two leagues in Major League Baseball (MLB) was divided into two divisions to accommodate the addition of extra teams to each league. During these years, the two divisional champions in each league played against each other in an initial playoff round to determine which teams would go on to represent the American and National Leagues in the World Series.

 In 1994, due to more expansion, each league was reorganized into a three-divisional structure. This led to the situation of having two sets of three divisional winners. In order to maintain a tidy playoff system, each league began sending a “wild card team” into the first of three rounds of playoffs. This wild card team would simply be the best of the three second place finishers and would join the three divisional winners in the first round of games.

 Ever since 1993, when major league baseball leaked its plan to use wild card teams, controversy and debate has ensued among baseball fans, media commentators, and team officials. In a 1993 article in Baseball America, Philadelphia Phillies president Bill Giles summarized his views of the new structure as “It should make September and the post season more interesting” (Hagan 1993, p. S5). However, in an earlier article in USA Today’s Baseball Weekly, Hall of Famer Johnny Bench exclaimed that “Someday, we’ll have a team that wins the World Series without winning their own division” (Lawes, 1992, p. 3). As recently as 2006, Pulitzer Prize nominee Rick Hummel complained in a column in the St. Louis Post-Dispatch that “Already baseball has been burned (some would say scarred) by the fact that three of the past four World Series winners have been wild card teams…. Baseball needs to make it much harder for a wild-card team” (Hummell 2006, p. C6).

 To understand the three division playoff system, consider the final American League standings from 2006 in Table 1 (obtained from wikipedia). As divisional champs, the New York Yankees, Minnesota Twins, and Oakland Athletics advanced. Among the second place teams, the Detroit Tigers had the highest winning percentage and hence was labeled as the wild card team. In the playoffs, the wild card Tigers defeated the Yankees and Athletics to advance to the World Series, where they lost to a division-winning St. Louis Cardinals team out of the National League. (It is interesting to note that the Cardinals only had the 5th-best record in the National league, but as division-winners advanced to the playoffs over two teams with better records.)

|  |
| --- |
| **TABLE 1:** Final American League Standings from 2006 |
| **East Division** |
|

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Team | W | L | WL% | GB |
| New York | 97 | 65 | .599 | --- |
| Toronto | 87 | 75 | .537 | 10.0 |
| Boston | 86 | 76 | .531 | 11.0 |
| Baltimore | 70 | 92 | .432 | 27.0 |
| Tampa Bay | 61 | 101 | .377 | 36.0 |

 |
| **Central Division** |
|

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Team | W | L | WL% | GB |
| Minnesota | 96 | 66 | .593 | --- |
| Detroit | 95 | 67 | .586 | 1.0 |
| Chicago | 90 | 72 | .556 | 6.0 |
| Cleveland | 78 | 84 | .481 | 18.0 |
| Kansas City | 62 | 100 | .383 | 34.0 |

 |
| **West Division** |
|

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Team | W | L | WL% | GB |
| Oakland | 93 | 69 | .574 | --- |
| Los Angeles | 89 | 73 | .549 | 4.0 |
| Texas | 80 | 82 | .494 | 13.0 |
| Seattle | 78 | 84 | .481 | 15.0 |

 |

 The aim of this article is to use elementary (college-level) probability to offer support to having wild card teams competing in baseball’s playoffs. To generate this support, we will apply a simple probabilistic model to the final MLB regular season standings. Using this model, we will show that the wild card team should typically be better than at least one of the division winners, just as in 2006. If there were no wild card teams, we will show that in 72% of seasons, a team would be left out of the playoffs while a team with a lower winning percentage would advance.

**2. The Probabilistic Model**

 Our probabilistic model is very simple. We will explain the model using the American League, and only state the results for the National League. To understand the model, allow the team names to be replaced with team ranks within the league. Therefore, the American League is comprised of teams ranked 1 through 14. With this ranking system, Team 1 in 2006 would be the New York Yankees, as they had the best record in the league. Minnesota would then be Team 2, and so on. In a different year, Team 1 might correspond to a different team, if some other team had the best record.

 The fundamental and, in fact, the only assumption that we make in our model is that the ranks of the 14 teams for any given season are randomly distributed across the three divisions. To better understand this assumption and the probabilistic results that follow, assume one holds a set of 14 balls with five inscribed East, five inscribed Central, and the remaining four inscribed West. To randomly assign a division to Team 1, the best team, a ball is chosen at random from the 14. Continuing this process for teams 2 through 14, without replacing the chosen balls, completes the assignment.

 Our probabilistic model considers two random variables. These are the rank of the wild card team, denoted *WC*, and the rank of the worst divisional winner, denoted *WDW*. We will show that the expected value *E(WC)* of the rank of the wild card team is smaller than the expected value *E(WDW)* of the rank of the worst divisional winner. In fact, the model suggests that, in over 5% of seasons, the worst divisional winner will rank in the bottom half of the league.

 Since there are three divisions, it is not hard to see that the wild card team (the best of the second place teams) will always have rank 2, 3, or 4. Similarly, the worst divisional winner can only have a rank from 3 through 11. In order to calculate the expected rank of the wild card team, one needs to apply the statistical notion of a discrete expected value:

*E(WC) = 2\*Pr(WC=2) + 3\*Pr(WC=3) + 4\*Pr(WC=4),*

where *Pr(X)* represents the probability that some event *X* occurs.

 Let’s first think about *Pr(WC=2)*. The only way to have *WC=2* is for Team 1 and Team 2 to be in the same division. We will use the multiplication rule for probability:

 *Pr(WC=2) = Pr(Team 2 is in East given Team 1 is in East) \* Pr(Team 1 is in East)*

 *+ Pr(Team 2 is in Central given Team 1 is in Central) \* Pr(Team 1 is in Central)*

 *+ Pr(Team 2 is in West given Team 1 is in West) \* Pr(Team 1 is in West).*

It is simple to see that *Pr(Team 1 is in East) = 5/14*, since there are five teams in the East and none of the 14 teams’ rankings are yet assigned. Then *Pr(Team 2 is in East given Team 1 is in East) = 4/13,* since there are only four East division slots left and 13 remaining teams. In this way,

*Pr(WC=2) = (4/13)(5/14) + (4/13)(5/14) + (3/13)(4/14) = 0.2857.*

Already, we have a very interesting result! Under this model, more than 28% of the time, the wild card team will be the second best team in the American League. Thus, without the wild card team, in about one year out of four the second-best team in the league would not advance to the playoffs.

 Now consider *Pr(WC=3).* In order to have *WC=3*, two things are necessary: Team 1 and Team 2 must be in different divisions and Team 3 must be in the same division as one of the others. In order to calculate all of the possible combinations, let *Q(X,Y)* represent the probability that Team 1 is in division *X*, Team 2 is in division *Y*, and Team 3 is in either division *X* or *Y*. If *E*, *C*, and *W* are the possible outcomes of *X* and *Y*, then

*Pr(WC=3) = Q(E,C) + Q(C,E) + Q(E,W) + Q(W,E) + Q(C,W) + Q(W,C).*

Now, we have

*Q(E,C) = Pr(Team 3 is in East or Central given Team 1 is in East and Team 2 is in Central)*

*\* Pr(Team 2 is in Central given Team 1 is in East) \* Pr(Team 1 is in East).*

With reasoning very similar to that used earlier, *Q(E,C)* can be seen to equal (8/12)(5/13)(5/14) and

*Pr(WC=3) = (8/12)(5/13)(5/14) + (8/12)(5/13)(5/14) + (7/12)(4/13)(5/14)*

 *+ (7/12)(5/13)(4/14) + (7/12)(4/13)(5/14) + (7/12)(5/13)(4/14)*

*= 0.4396.*

 Although we will not show the details, we can use similar reasoning to show that *PR(WC=4) = 0.2747*. We now see that *E(WC) = 2(0.2857) + 3(0.4396) + 4(0.2747) = 2.989*. Since this expected rank is less than 3 and the worst divisional winner must have rank 3 or larger, our model suggests that the wild card team is, on average, better than the worst divisional winner!

 It is also possible to find the probability that the wild card team is better than the worst divisional winner. Since this happens exactly when the wild card team has rank 2 or 3,

*Pr(WC<WDW) = Pr(WC=2) + Pr(WC=3) = 0.2857 + 0.4396 = 0.7253.*

So the model suggests that wild card team will be better than the worst divisional winner about 72.53% of the time.

**3. The Worst Divisional Winner**

 These results show just how successful we can typically expect the wild card team to be. Now, how good can we expect the worst divisional winner to be? Clearly, the best possible rank that the worst divisional winner can attain is 3. At the opposite extreme, if the four worst teams all play in the West division, the WDW will have rank 11. In fact, in 1994 when the season ended early due to a player’s strike, the Texas Rangers were the first place team in the West and were also ranked 11th out of the 14 teams based on winning percentage. So this worst case scenario is not that farfetched.

 The expected rank of the worst divisional winner is given by:
$$E\left(WDW\right)= \sum\_{n=3}^{11}n ∙Pr\left(WDW = n\right).$$

The probabilities in this sum are once again found by conditioning on which division contains Team *n*. Specifically,

*Pr(WDW=n) = Pr(WDW=n and Team n wins the East)*

 *+ Pr(WDW=n and Team n wins the Central)*

 *+ Pr(WDW=n and Team n wins the West).*

By using successive conditional probabilities, the following type of substitution is used in the above expression:

*Pr(WDW=n and Team n wins the East)*

 *= Pr(WDW=n and Team wins the East and Team n is in the East)*

 *= Pr(WDW=n given Team n wins the East and Team n is in the East)*

 *\* Pr(Team n wins the East given Team n is in the East)*

*\* Pr(Team n is in the East).*

 To apply this formula, consider for example the case *n=5*. As we have already seen, *Pr(Team 5 is in the East) = 5/14*. For Team 5 to be the worst divisional winner, two things must occur. Team 5 must outplay the other four teams in the East and Teams 1-4 must not all be in the same division. Each of these described occurrences effectively suggests one of the conditional probabilities in the formula above.

 Assuming that Team 5 is an East division team, there are $\left(\genfrac{}{}{0pt}{}{9}{4}\right)$ ways to allocate teams whose rank is larger than 5 into the East division. For each of these combinations, the probability of the four teams winding up in the East is (4/13)(3/12)(2/11)(1/10). The resulting probability of any four lower ranked teams being placed in the East with Team 5 becomes
$$Pr(Team 5 wins the East given Team 5 is in the East) = \left(\genfrac{}{}{0pt}{}{9}{4}\right)(4/13) (3/12) (2/11)(1/10) = 0.1762.$$

Now we must find the probability that the top four teams are not all in the same division. The probability that they ARE in either all in the Central or all in the West division is given by (5/9)(4/8)(3/7)(2/6) + (4/9)(3/8)(2/7)(1/6) = 0.0476. The first product is the probability that they are all in the Central, and the second is the probability that they are all in the West. So the probability that the top four teams ARE NOT in the same division is given by 1 – 0.0476 = 0.9524. This shows that

*Pr(WDW=5 given Team 5 wins the East and Team 5 is in the East) = 0.9524.*

Putting the three probabilities together,

*Pr(WDW=5 and Team 5 wins the East) = (0.9524)(0.1762)(5/14) = 0.0599.*

With some similar work, we can find that

*Pr(WDW=5) = 0.0599 + 0.0599 + 0.0799 = 0.1997.*

 Using the same method, we can calculate the values of *Pr(WDW=n)* in Table 2. It is interesting to note that the sum of the last four probabilities in Table 2 is 0.056. Hence, the worst divisional winner will fall in the bottom half of the league 5.6% of the time.

|  |
| --- |
| **TABLE 2:** The Distribution of the Ranking of the Worst Divisional Winner |
|

|  |  |
| --- | --- |
| *n* | *Pr(WDW=n)* |
| 3 | 0.2747 |
| 4 | 0.2747 |
| 5 | 0.1998 |
| 6 | 0.1249 |
| 7 | 0.0699 |
| 8 | 0.0350 |
| 9 | 0.0150 |
| 10 | 0.0050 |
| 11 | 0.0010 |

 |

Using the formula for the expected value, we have $E\left(WDW\right)= \sum\_{n=3}^{11}n ∙Pr(WDW = n)= 4.6354$. This implies that *E(WDW – WC) = E(WDW) - E(WC) = 1.6474*. That is, we expect that, on average, the wild card team will rank 1.6474 spots ahead of the worst division winner in the overall league standings.

**4. Comparing the Model to Reality**

 By reviewing the standings of the American League from 1994 through 2010, we can compare the average results of the probabilistic model to what actually occurred during those 17 seasons. These comparisons are shown in Table 3.

**TABLE 3:** Model Estimates and Actual American League Values

 *E(WC) E(WDW) Pr(WC<WDW)*

Model suggests: 2.989 4.635 0.725

Actual outcomes: 2.853 4.441 0.765

**5. The National League**

 The techniques we used here can of course also be applied to the National League. This is different from the American League in that there are 16 teams, with five in the East and West divisions and six in the Central division. Readers are encouraged to apply the steps and techniques given above to this slightly different arrangement, and should find the following results:

*E(WC) = 2.9762*, *E(WDW) = 4.7164*, and Pr*(WC<WDW) = 0.7321*.

Also, the worst divisional winner should fall in the bottom half of the National League about 2.9% of the time. In reality, from 1994 through 2010 the average rank of the wild card team in the National League was 3.088, the average rank of the worst divisional winner was 4.471, and the wild card team was better than the worst divisional winner with probability 0.588.

**6. Conclusions and Further Reading**

 While the objections noted by many baseball enthusiasts are understandable, these results suggest that the worst divisional winner is typically the weakest link in the three division playoffs structure. In over 72% of seasons, the wild card team will be better than at least one of the other teams in the playoffs. This finding seems to contradict Johnny Bench’s belief that wild card playoffs elevates mediocrity. Rather, it seems that the average quality of the playoff teams increases when the wild card team is included. If the purpose of a playoff structure is to allow the best teams to compete for the championship, a system with a wild card team is more desirable than one without.

 Readers interested in learning more about applying simple probability models to baseball applications are encouraged to read either of the following two excellent books:

Jim Albert, *Teaching Statistics Using Baseball* (The Mathematical Association of America, 2003),

Ken Ross, *A Mathematician at the Ballpark* (Plume, The Penguin Group, 2007).

**References**

Hagan, P. (1993), “Owners Respond to Crisis, Discuss Sweeping Changes,” *Baseball America,* 13(5), p. S5.

Hummel, R. (2006), “Around the Horn,” *St. Louis Post-Dispatch*, June 4, p. C6.

Lawes, R. (1992), “Playoff Expansion: A New View,” *USA Today’s Baseball Weekly*, 2(30), p. 3.