

# Presidents, Baseball, and Wins above Expectations: What Can Sabermetrics Tell Us about Presidential Success?

## Why Ronald Reagan is like Bobby Cox and Lyndon Johnson is like Joe Torre

Manuel P. Teodoro, *Texas A&M University*

Jon R. Bond, *Texas A&M University*

### ABSTRACT

Presidential scholars and baseball writers debate who were the greatest. While baseball analysis evolved from qualitative impressions of “experts” to rigorous, data-driven “sabermetrics,” analysis of presidential greatness continues to rely on “old-school” reputational rankings based on surveys of scholars’ qualitative assessments. Presidential-congressional relations and baseball are all about winning, but what fans (of sports and politics) find most intriguing is Wins Above Expectations (WAE)—did the team do better or worse than expected? This paper adapts the Pythagorean Expectations (PE) formula developed to analyze baseball to assess legislative success of presidents from Eisenhower to Obama. A parsimonious regression model and the PE formula predict annual success rates with 90% accuracy. The estimates of WAE from the two approaches, however, are uncorrelated. Regression analysis does not identify any president who systematically exceeded expectations, but sabermetric analysis indicates that Republican presidents outperform Democrats. Neither approach correlates with recent presidential greatness rankings.

“You can observe a lot by watching.”  
—Yogi Berra (1980)

Presidential scholars and baseball writers have spilled barrels of ink on a perennial question: Who is the G.O.A.T. (that’s Greatest Of All Time)? With few objective standards to guide comparisons across eras, fans of both politics and baseball have assessed greatness impressionistically, relying on intuition and the meager statistics available to form evaluations. But over the past 35 years, baseball analysis has evolved into a rigorous, data-driven endeavor.

Long scorned by “old-school” baseball writers, the sabermetric approach is now recognized—albeit grudgingly by some—as a powerful way to assess teams, players, and managers objectively and with precision across time. Valid and reliable metrics now inform baseball fans’ debates.

Alas, developing metrics to assess the performance of US presidents lags far behind. Our understanding of presidential “greatness” continues to rely on “old-school” reputational rankings based on surveys of scholars’ impressions that average an unknown mix of subjective and objective criteria, and on in-depth case studies that analyze a small number of important, but unrepresentative, cases of presidential successes and failures.

Taking some pages from the baseball playbook, this article adapts a sabermetric approach to assess the legislative success of presidents from Eisenhower to Obama. Bill James (1982)

Manuel P. Teodoro is an associate professor of political science at Texas A&M University in College Station. He can be reached at [mteodoro@tamu.edu](mailto:mteodoro@tamu.edu).

Jon R. Bond is a professor of political science at Texas A&M University in College Station. He can be reached at [jonbond@tamu.edu](mailto:jonbond@tamu.edu).

developed the Pythagorean Expectations (PE) formula to analyze success in baseball. The PE formula estimates how many games a team should have won based only on runs scored and runs allowed.

Following the sage advice of our epigraphic Yogi-ism, we observe some parallels between presidential-congressional relations and baseball. Like baseball, presidential-congressional relations are all about winning. What a baseball manager needs from his players are runs: hitters must score runs; fielders must prevent them. What a president needs from members of Congress are votes: “votes for and against bills, votes for and against amendments,” votes to sustain vetoes, and “votes (every century or so)” opposing impeachment (King 1983, 247).

But important as winning is, what sports geeks and political scientists find most intriguing is Wins Above Expectations (WAE)—did the team do better or worse than expected? If a team wins more games than expected, was it due to serendipitous

these findings suggest that the three approaches capture different aspects of the underlying data generation process.

#### POLITICAL SCIENCE ESTIMATES OF PRESIDENTIAL WINS ABOVE EXPECTATIONS

Richard Neustadt (1960, chap. 4) argued that the president’s “professional reputation,” or political skill, was the key to success. We can state this as the Political Skills Hypothesis: *ceteris paribus*, presidents who are highly skilled at bargaining and persuasion should win more than expected, while less skilled presidents should win less. The primary challenge to testing the effects of skill is identifying valid and reliable quantitative measures of political skills that are comparable across time.

Evidence supporting the skills hypothesis comes mainly from case studies. This evidence is mixed at best, and generalizing from selected (and unrepresentative) cases is problematic. Even a multiple-cases design that uses common standards to assess the

*If a roll call vote on which the president expresses a position is analogous to a game, and votes for and against the president are analogous to runs scored and runs allowed, can the PE formula predict presidential success rates over the course of a season (year) as accurately as it does for professional baseball teams?*

circumstances, simple “luck,” or did the manager’s leadership produce wins when the team should have lost? If a roll call vote on which the president expresses a position is analogous to a game, and votes for and against the president are analogous to runs scored and runs allowed, can the PE formula predict presidential success rates over the course of a season (year) as accurately as it does for professional baseball teams? If so, how well do WAEs from PE correlate with the closest political scientists have come to quantitative estimates of presidents’ WAEs—i.e., using regression residuals to identify presidents who won more or less than expected given conditions captured in the models? Which presidents, if any, stand out as uncommonly successful or unsuccessful at winning votes in Congress? How do quantitative estimates compare to the received wisdom of scholarly surveys regarding which presidents were great legislative leaders?

The article proceeds as follows. We begin with a review and update of political scientists’ attempts to estimate presidents’ legislative WAEs with regression analysis. Next, we adapt the Sabermetric PE formula to predict presidential success rates based only on votes for and against, and present WAE relative to this parsimonious, atheoretical baseline. We find that both approaches predict presidents’ legislative success rates with high accuracy across all years and presidents. The WAEs from the two approaches, however, are uncorrelated: multivariate regression models yield no evidence of systematic variation in presidential WAE, but the sabermetric analysis indicates that some presidents were markedly more successful than others, and that Republican presidents generally outperform Democrats. In homage to the old-school-baseball-writers-versus-sabermetricians rivalry, we look at recent surveys of presidential scholars, finding that scholars’ assessments of presidential legislative skill are uncorrelated with our sabermetric evaluation. Taken together,

performance of several presidents (Kellerman 1984), and studies that develop clever quantitative measures of aspects of political skills<sup>3</sup> are plagued by validity issues, ambiguities in interpreting findings, and generalizability (Bond and Fleisher 1990; Fleisher, Bond and Wood 2008). Nonetheless, reviews of the case-study literature found consensus that Lyndon Johnson and Ronald Reagan were the most highly skilled legislative leaders since FDR, and that Richard Nixon and Jimmy Carter were the least skilled (Bond and Fleisher 1990, 198–204; Edwards 2009, chap. 4).

Another controversial approach seeks to finesse the intractable measurement problem and test the Political Skills Hypothesis indirectly. This approach uses regression models to estimate the effects of contextual variables (e.g., party, public approval, and polarization) on presidential success on roll call votes in Congress.<sup>4</sup> These models establish a common baseline against which to compare presidents, and the errors can be interpreted as estimates of Wins Above Expectations—whether the president won more or less often than should be expected given the political context. Residuals from regression analysis, of course, are not a measure of skill. The error term contains everything not explained by variables in the model, and residuals should represent random error. Nevertheless, studies using this approach are in effect following advice from *The Sabermetric Manifesto*: “a claim which cannot be directly tested can be evaluated by studying the conclusions which would follow” (Grabner n.d.).

What might we expect to see if the Skills Hypothesis were true? If political skill is as important as Neustadtian theory suggests, then the effects of this omitted variable should show up as non-random variance in the errors. To be clear, even if residuals of the consensus candidates for especially skilled (Johnson and Reagan) or unskilled (Nixon and Carter) presidents are unusually large, we cannot interpret that as evidence of skill. But repeated failures to reject the null-hypothesis that the errors are random

should raise doubts about leadership skill as a *systematic* explanation of presidential success on roll call votes.

Studies using different model specifications analyzing different periods were unable to reject the null hypothesis that the residuals are random error (Bond and Fleisher 1990, chap. 8; Cohen, Bond, and Fleisher 2013; Fleisher and Bond 1983, 1992; Fleisher, Bond, and Wood 2008). Even the tried-and-flawed “inter-ocular test”—eyeball the plot, and if a pattern hits you between the eyes, it’s significant—failed to reveal expected patterns. An updated analysis of annual presidential success rates from 1953–2014 explains 89% of the variance in the House and 72% in the Senate with three variables (party control, public approval, and party polarization) plus interactions.<sup>5</sup> Figure 1 plots studentized residuals from these regressions.<sup>6</sup> Diagnostics found no evidence of heteroscedasticity, autocorrelation, specification error, or omitted variables. Moreover, the “inter-ocular test” again failed to find evidence that Johnson and Reagan won significantly more than should be expected, or that Nixon and Carter won less. Perhaps sabermetrics can provide additional leverage on estimating presidents’ wins above expectations.

### SABERMETRIC ESTIMATES OF PRESIDENTIAL WINS ABOVE EXPECTATIONS

Political scientists chase the god of statistical significance (anything more likely than .05 must be rejected) in search of theory. Sabermetricians chase the god of big (really big) data in search of accurate predictions. With big data, statistical significance and theory fade into irrelevance as ever more accurate of predictions reveal empirical truth. If asked about theory, a sabermetrician might reference Alfonso Bedoya in *Treasure of the Sierra Madre*: “Theory? We ain’t got no theory. We don’t need no . . . stinkin’ theory!”<sup>7</sup> So let’s take off our shoes and sox and see what shoeless empiricism reveals.

#### Pythagorean Expectations

Bill James’ (1982) Pythagorean Expectations formula is perhaps the foundational element of sabermetrics. It predicts Major League

Baseball teams’ winning percentages with over 90% accuracy based solely on runs scored and runs allowed:

$$\text{Win Percentage} = \frac{\text{runs scored}^\gamma}{\text{runs scored}^\gamma + \text{runs allowed}^\gamma}$$

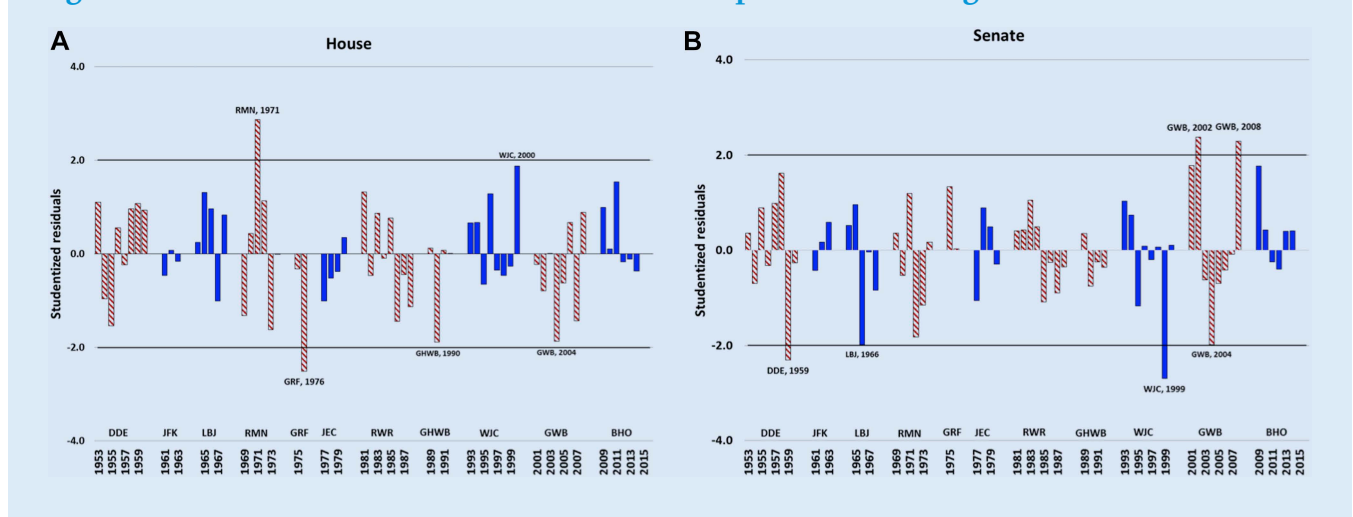
The central insight of James’ PE formula is that runs scored and runs allowed reveal information about a team’s performance beyond whether it won or lost any particular game. Although the binary outcome may be the primary concern—Does the team win the game? Does the president win the roll call?—binary outcomes are the product of some continuous underlying data generating process. Each run scored or allowed implies a story of successes or failures that culminated in a win or loss, and so runs scored and runs allowed are better indicators of actual team performance—Are the hitters scoring? Are the pitchers and fielders making outs?—than is simple winning percentage. Over the course of a 162 game season for 30 Major League franchises, winning percentage and PE converge quite closely. In James’ original formulation, the value of  $\gamma$  was set at 2 for the sake of simplicity and because it fit the data reasonably well.<sup>8</sup> Since then, the PE formula for major league baseball has been refined to arrive at  $\gamma = 1.82$  as the exponent that generates the most accurate predictions (Miller 2007).

PE does not perfectly predict actual winning percentages, and sabermetricians hotly debate the meaning of deviations from PE, or Wins Above Expectations (WAE). Put simply, WAE is a measure of games that a team won that it should have lost (positive WAE), or lost that it should have won (negative WAE), over the course of a season. For many, WAE is merely noise that reflects the irreducible randomness of a game played by humans under varying conditions. For others, WAE indicates the quality of relief pitching, the ability to hit or pitch “in the clutch” (i.e., in high-leverage situations), or—critically for our purposes—the skill of a team’s manager.<sup>9</sup>

#### WAE as a Measure of Managerial Skill

The intuition behind WAE as a measure of managerial skill is that, over the course of a season, a baseball team’s PE is mostly a function of the talent on its roster. A team full of elite athletes

Figure 1  
Regression Estimates of Presidents’ Wins Above Expectation in Congress



in their primes should score plenty of runs and allow few, with or without a great manager. Just so, a team with slower, weaker, or injury-prone players probably will score few runs and allow many, regardless of what its manager does. WAE, then, is argued to measure how effectively a manager deploys the talent at his disposal in close games.

Two championship seasons from the career of Hall of Fame manager Sparky Anderson—the 1976 Cincinnati Reds and 1984 Detroit Tigers—are instructive. Fresh off a World Series championship in 1975, the 1976 Reds were a juggernaut—the team’s roster boasted three future Hall of Famers (plus Pete Rose, whose off-field transgressions have kept him from enshrinement). Dubbed the

majorities in Congress is like the manager of a team full of poor hitters and soft-tossing pitchers who struggle to throw strikes; under such conditions, any roll call wins at all might be evidence of the president’s political skill, even though the actual winning percentage is low.

Just as James’ PE formula recognized that runs scored and allowed convey information about team performance that is not reflected in winning percentage, votes won and lost on presidential roll calls may contain information about presidential performance beyond an annual winning percentage. Each vote for or against the president carries information about legislative politics, including party control, party cohesion, and public approval.

*We observe a number of parallels between presidential-congressional relations and baseball. The president is analogous to a manager. A roll call is analogous to a game that the president’s team plays. Votes supporting and opposing the president are analogous to runs scored and allowed in each “game.” A year is analogous to a season. A president whose party holds large majorities in both houses of Congress is akin to a manager with an all-star roster; we would expect such a president to win most roll calls without extraordinary effort.*

“Big Red Machine” in the press, the team cruised to a 102–60 record, ten games ahead of the second-place Dodgers, and swept the playoffs en route to a second consecutive World Series title. Although baseball fans revere the 1976 Reds as one of baseball’s all-time greatest teams, its 102 wins were actually less than its Pythagorean Expectation of 103–59, giving Sparky Anderson a -1.0 WAE (or -0.62 WAE%) for the season.

By contrast, the 1984 Tigers were building on a solid but unspectacular performance in 1983 with a roster that included no Hall of Famers.<sup>10</sup> Yet the 1984 Tigers led the American League all season, swept the Royals to win the American League pennant, and beat the Padres in five games to win the World Series. The 1984 Tigers’ 104–58 record exceeded their Pythagorean Expectation by five games, giving Anderson an extraordinary 5.0 WAE and 3.09 WAE%. Not coincidentally, Anderson was named the 1984 American League Manager of the Year.

The point of the illustration is simple but enlightening. The 1976 Reds were expected to crush the National League with their talent-laden roster, so the team’s excellent record and championship offer little evidence of Anderson’s managerial prowess; the 1984 Tigers were expected to be merely good, so the team’s dominant performance that year suggests that Anderson’s management contributed to its success.

### Presidential PE in Congress

Can the PE formula be adapted to predict presidential success in Congress? We observe a number of parallels between presidential-congressional relations and baseball. The president is analogous to a manager. A roll call is analogous to a game that the president’s team plays. Votes supporting and opposing the president are analogous to runs scored and allowed in each “game.” A year is analogous to a season. A president whose party holds large majorities in both houses of Congress is akin to a manager with an all-star roster; we would expect such a president to win most roll calls without extraordinary effort. A president who faces hostile

The typical regression analysis considers each of these variables separately on the right-hand side of its equation with specific causal theories in mind. The PE approach is unapologetically atheoretical, aggregating votes over a year as a way of approximating the overall strength of a president’s legislative sway apart from any one roll call, just as the baseball PE measures the strength of a team’s roster apart from any one game. Adapted to presidential-congressional politics, the formula for Pythagorean Expected roll call winning percentage (PE%) is:

$$PE\% = \frac{\text{votes for}^y}{\text{votes for}^y + \text{votes against}^y}$$

The difference between actual winning percentage and PE% is percentage Wins Above Expectations (WAE%). At the end of a year, a positive value indicates that the president won more roll calls than expected, while a negative value indicates more losses than expected. This percentage multiplied by the number of roll calls on which the president took a position yields Wins Above Expectations (WAE). Just as some baseball analysts use WAE to assess managerial skill, WAE might also yield information about presidents’ skill with legislative politics. As with baseball managers, WAE% offers a useful way to think about presidents’ annual performance, while total WAE rewards performance across several years and so offers some perspective on career legislative accomplishments.

The need to use WAE% rather than WAE highlights two important differences between baseball games and congressional roll calls.<sup>11</sup> First, whether enjoying a winning season or suffering through a miserable one, every Major League team plays 162 games. “You win a few, you lose a few, some get rained out,” Satchel Paige once observed, “but you got to dress for all of them.” Presidents win a few and lose a few, but unlike ballplayers, presidents don’t have to dress for all of them. Instead, presidents get to pick which “games” to play when they choose whether to take a position on a particular roll call. Being able to pick their votes raises the

possibility that a president may seek to manipulate their success rate and WAE by sitting out a sure loser or stepping up to endorse a sure winner. Although a president may occasionally engage in such posturing, Peterson (1990) presents evidence that presidential positions are generally sincere. And Covington (1987b) finds evidence that the president can occasionally increase the chances that a roll call will go their way by “staying private.” Such cases are rare, but these “wins” do not appear in the success rate because the president remained in the clubhouse. Nevertheless, analysis of presidential “dressing for the game” reveals no evidence that presidents’ position taking systematically affects their actual legislative success rate or WAE (see appendix A).

Second, for analytical purposes, in baseball each run scored is an independent event: the probability that a team will score or allow a run in a given game is essentially unrelated to the runs scored and allowed in the previous and subsequent games. One reason for the success of sabermetrics is that baseball scoring satisfies the independent, identically distributed (IID) assumption. The validity of the IID assumption for congressional roll calls is more questionable: the probability that a president will lobby effectively on legislation may depend on their past success or future legislative agenda. In this analysis, however, such a violation of the IID assumption might be analytically useful if the interdependence of roll calls reveals something about Neustadt’s Political Skills hypothesis. Presidents may gain or lose momentum during a given session of Congress as a function of their legislative skill. With this caveat in mind, we now apply a sabermetric framework to presidential politics.

### Calculating Presidential PE

Following standard sabermetric procedures, we used the presidential roll call vote data described earlier to calculate PE WAE% and WAE for presidents. Votes for and against the president’s position were totaled for roll calls on which the president took a position for each year from 1953–2014 ( $n=4231$  House votes, 4356 Senate votes).<sup>12</sup> Using these totals, we calculated PE for each year with  $\gamma$  set to 2 using James’ (1982) original formulation. The result was a striking .934 Pearson correlation between PE% and actual winning percentage in the House and .907 in the Senate. These impressive correlations belied substantial year-to-year error, so we followed the usual sabermetric practice of minimizing squared differences between PE% and actual winning percentage. We applied a Generalized Reduced Gradient (GRG) iterative algorithm to the PE% formula and roll call voting data with an aim of minimizing squared differences, with  $\gamma$  constrained to be greater than zero and less than 100. The GRG algorithm is an efficient, robust method for optimizing nonlinear problems (Mantell and Lasdon 1978).<sup>13</sup> Optimization yields  $\gamma$  values of 3.44 in the House and 3.11 in the Senate. Thus, the PE% formulae for presidents are:

$$\text{House PE\%} = \frac{\text{votes for}^{3.44}}{\text{votes for}^{3.44} + \text{votes against}^{3.44}}$$

$$\text{Senate PE\%} = \frac{\text{votes for}^{3.11}}{\text{votes for}^{3.11} + \text{votes against}^{3.11}}$$

### RESULTS OF SABERMETRIC ANALYSIS OF PRESIDENTIAL WINS ABOVE EXPECTATIONS

These formulas generate PE% that correlate .939 with actual winning percentages in the House and .829 in the Senate.<sup>14</sup> The slightly

higher  $\gamma$  value for the House reflects the wider average winning margins in that chamber.<sup>15</sup> Figure 2 shows annual WAE% for House and Senate, standardized to facilitate comparison to the studentized residuals from the regression analysis.

Although the regression models and the PE formula predict presidential success rates with similarly high levels of accuracy, the WAE% from the two approaches are uncorrelated ( $r^2 = 0.001$  and House and 0.078 in the Senate). The patterns of PE WAE% are quite different from those estimated by studentized residuals. The studentized residuals indicate that each president exceeded expectations in some years and fell short in others more or less at random in both chambers—exactly as expected for regression models. The PE estimates, however, indicate that presidents generally were consistently above or below expectations in both chambers across the years of their administrations. In particular, Reagan was considerably above the mean in both the House and Senate during all eight years. Clinton was below the mean for all but one or two years of his presidency, and well below in the Senate. Obama exceeded expectations in the House during his first two years, but falls far below the mean in the next four years.

Moreover, the PE approach produces a strong partisan pattern: Republican presidents outperform Democrats. Ike is in last place, but otherwise the top WAEs are all GOP. Neither regression analysis nor case studies and scholar surveys find evidence of a party advantage.

### Sabermetric Analysis of Neustadt’s Skills Hypothesis

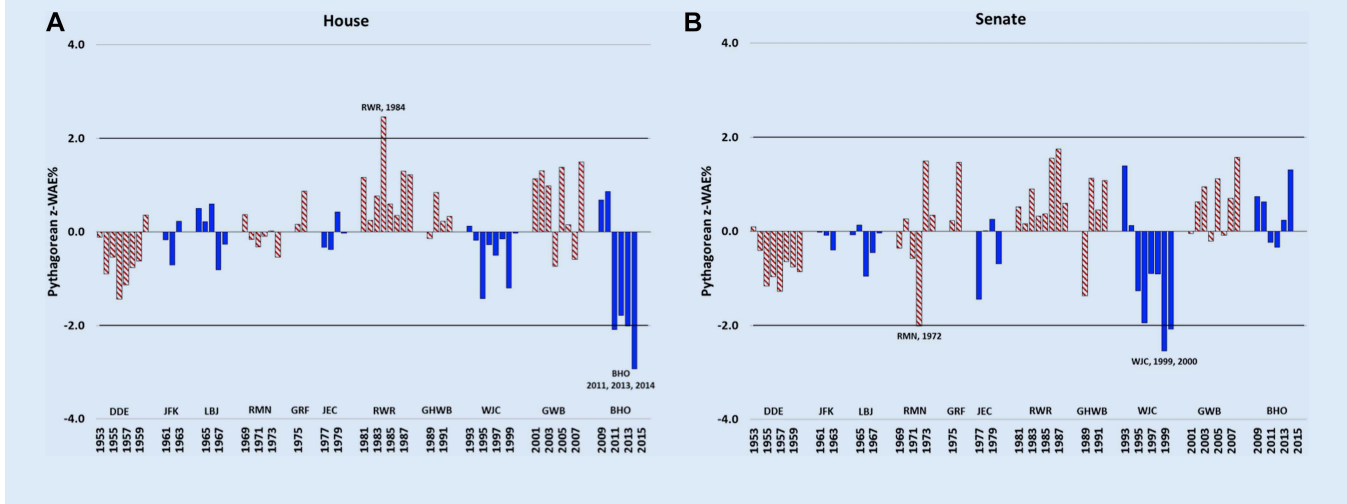
How does Neustadt’s Political Skills Hypothesis fare with sabermetric WAE? Do presidents who are renowned for legislative skill generate impressive WAE? And does WAE% in the House correlate with WAE% in the Senate, as the skills hypothesis suggests? To address these questions, we used PE% to calculate annual WAE% and WAE for each chamber. We also calculated each president’s “career average” WAE% and WAE by summing their annual WAE across chambers. Table 1 reports these results<sup>16</sup> ordered by career WAE.<sup>17</sup>

Sabermetric analysis offers mixed support about the political skill of the reputed legislative giants Lyndon Johnson and Ronald Reagan. Johnson’s career statistics place him first among Democrats, but in the middle of the lineup overall, with House ( $hWAE\% = 2.9$ ,  $hWAE=11.4$ ) and Senate ( $sWAE\%=1.6$ ,  $sWAE=7.9$ ) averages below the mean among post-WWII presidents. In contrast, Reagan’s career statistics make him by far the most impressive legislative operator in the sample, ranking second in career WAE% and an easy first place in WAE in both houses. Reagan’s career  $hWAE$  of 82.5 and  $sWAE$  of 51.7 are more than double his next closest rival’s.

On the low end of the reputational distribution are Richard Nixon and Jimmy Carter. Here, sabermetric assessment diverges from the scholarly consensus. Their WAEs are near the middle of lineup—Nixon ranks #5 (just ahead of LBJ), with Carter just behind him. In sabermetric terms, it is more accurate to describe their legislative records as “mediocre” rather than “bad.”

Finally, Neustadt does not suggest that influencing the House and Senate requires different skills, so we should observe similar patterns in both chambers. Sabermetrics offer some support for this expectation. The Spearman rank-order correlation of career House and Senate WAE% is .69 ( $p=.02$ ), indicating that the rankings are broadly similar in both chambers.

Figure 2  
Pythagorean Presidential Wins Above Expectations



Sabermetric Analysis vs. Scholars’ Reputational Rankings

For comparison to a non-quantitative assessment, table 1 also reports each president’s ranking on their “Relations with Congress” from the 2010 Siena Presidential Expert Poll of presidential scholars, and a 2015 poll of political scientists who are members of the Presidency and Executive Politics Section of APSA (Rottinghaus and Vaughn 2015).

Both polls rank Johnson the top legislative leader and Carter last, but there is less consensus about the rankings in between (Spearman’s  $\rho = .591$ ,  $p = .06$ ). The ranking based on sabermetric career WAE, however, is uncorrelated with either poll ( $\rho = -.009$  with the

Siena poll and  $\rho = -.036$  with the APSA poll). The lack of correlation implies that presidential experts apply their own adjustments to presidents’ legislative records to account for context, but evidently not the same kinds of adjustments as those implied by PE. As with sabermetricians and traditional sports writers, quantitative analytics and expert qualitative assessments lead to different conclusions about presidents’ legislative prowess.

How do we reconcile these differences between presidents’ historic reputations and their sabermetric profiles? Baseball again offers an enlightening metaphor: in sabermetric terms, the managerial careers of Joe Torre and Bobby Cox parallel the legislative

Table 1  
Sabermetric Wins Above Expectations

President	House		Senate		Congress Overall			Scholar Rankings “Relations w/ Cong”	
	WAE%	WAE	WAE%	WAE	WAE%	WAE	Actual Win%	Siena 2010	APSA 2015
<i>Reagan</i>	12.0	82.5	9.0	51.7	10.5	134.2	56.8	7	3
<i>Bush43</i>	8.8	33.9	7.4	21.2	8.1	55.1	67.3	32	12
<i>Ford</i>	13.4	19.7	12.7	19.3	13.0	39.0	47.6	17	14
<i>Bush41</i>	6.1	24.9	5.4	13.2	5.7	38.1	45.4	23	9
<i>Nixon</i>	2.4	8.3	1.6	22.2	2.0	30.5	60.1	36	10
<b>Johnson</b>	<b>2.9</b>	<b>11.4</b>	<b>1.6</b>	<b>7.9</b>	<b>2.3</b>	<b>19.3</b>	<b>79.2</b>	<b>1</b>	<b>1</b>
<b>Carter</b>	<b>2.8</b>	<b>13.3</b>	<b>-1.0</b>	<b>1.5</b>	<b>0.9</b>	<b>14.8</b>	<b>73.1</b>	<b>39</b>	<b>18</b>
<b>Kennedy</b>	<b>1.6</b>	<b>2.4</b>	<b>1.4</b>	<b>4.8</b>	<b>1.5</b>	<b>7.2</b>	<b>82.3</b>	<b>13</b>	<b>11</b>
<b>Obama</b>	<b>-6.7</b>	<b>-34.8</b>	<b>5.9</b>	<b>19.0</b>	<b>-0.4</b>	<b>-15.9</b>	<b>66.7</b>	<b>18</b>	<b>13</b>
<b>Clinton</b>	<b>-0.4</b>	<b>-6.2</b>	<b>-5.4</b>	<b>-14.7</b>	<b>-2.9</b>	<b>-21.0</b>	<b>50.9</b>	<b>25</b>	<b>4</b>
<i>Eisenhower</i>	-2.0	-5.5	-3.3	-15.5	-2.6	-21.0	65.9	10	6
<b>Mean</b> (Across Years)	0.0	2.3	0.0	2.0	0.0	2.2	63.0		
<b>Mean</b> (Across Presidents)	3.7	13.6	3.2	11.9	3.5	25.5	63.2		

Entries listed in descending order of career Wins Above Expectations. Republicans in *italics*, Democrats in **bold**.

records of presidents Johnson and Reagan. Torre and Cox each managed major league teams over 29 years, both are widely considered all-time great managers, and both were inducted to the Hall of Fame in 2014. Over his managerial career, Torre won 2,326 games (.538 winning percentage), six American League pennants, and four World Series titles; Cox won 2,504 games (.556 winning percentage), five National League pennants, and one World Series. But their places on Darowski's all-time managers' WAE list are quite different:<sup>18</sup> Cox ranks 4<sup>th</sup> with a career WAE of 77.0, but Torre's respectable 23.3 WAE puts him far behind at 39<sup>th</sup>. A close look at their records makes the reason clear: Torre's most successful years came with the New York Yankees between 1996 and 2007, when his roster included some of the most fearsome players in baseball. Torre's Yankees teams won big, but they were supposed to win big based on their enormous payrolls and all-star laden rosters. Cox spent nearly his entire managerial career with the talented but less-than-awesome Atlanta Braves, winning consistently and routinely outperforming his PE.

Like Torre and Cox, Johnson and Reagan earned their reputations in different ways. Johnson's actual legislative winning percentages are high, but not much better than his high PE. Johnson won most of his legislative battles, including some extraordinary legislative "championships" with civil rights laws and Medicare. But like Torre's Yankees of the late-1990s and early-2000s, LBJ was *expected* to win with his strong partisan majorities in both houses and a booming economy. Reagan's actual winning percentages are relatively low (.420 House, .720 Senate), but like Bobby Cox, his record was consistently much higher than his PE, particularly in the House. Torre, Cox, Johnson and Reagan were all successful and probably deserving of their places in their respective Halls of Fame. But sabermetrics makes a stronger case for Cox and Reagan, while traditionalists would point to Torre's four World Series rings and Johnson's landmark legislative achievements in making their arguments for enshrinement.

## DISCUSSION

This analysis shows that the Pythagorean Expectations formula predicts presidential success rates about as accurately as well-specified regression models. What PE extracts from the sum of votes supporting and opposing the president, however, is different from the theoretical explanations of presidential success—the interaction of party control, public approval, and party polarization—included in our regression models. But unlike regression residuals, presidential WAE suggests that some presidents have been consistently above or below expectations. Presidential WAE also diverges markedly from scholars' collective, qualitative judgment of presidents' legislative efficacy as measured by widely cited surveys. Perhaps WAE is capturing something about crafting legislative coalitions, but as is the case with regression residuals and expert judgments, we don't know exactly what underlies the data generating process.

Sabermetrics may be a useful tool to analyze presidential success on roll call votes, but this paper barely scratches the surface. The more important contributions of sabermetrics to baseball are estimating players' Wins Above Replacement (WAR) to show how much each one contributes to his team's success. The parallel between members of the president's party and players, however, is much less clear than the parallel between votes and runs. At the very least, we think sabermetrics may offer a new, WAE cool direction for the study of presidents' participation in legislative politics.

## SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1049096516002778>

## ACKNOWLEDGMENTS

We are grateful to Carlisle Rainey, Guy Whitten and the participants in the Program in Scientific Political Methodology at Texas A&M University for helpful comments and suggestions. ■

## NOTES

1. Bill James defined "sabermetrics" as the "the search for objective knowledge about baseball." The name honors the Society for American Baseball Research (SABR).
2. Readers who don't want to chew the bubblegum are invited to skip to the presidential baseball cards in Appendix B.
3. These include Covington (1987a; 1987b; 1988a; 988b); Covington, Wrighton, and Kinney (1995); Fett (1994); Lockerbie and Borrelli (1989); and Sullivan (1988; 1990; 1991).
4. Winning roll call votes is only part of legislative success. While the floor vote is not the only important decision point in the legislative process, it is, nonetheless, a crucial decision that contains valid and reliable information that can be systematically observed. Members of Congress care about the votes they cast, because presidents, party leaders, interest groups, journalists, and voters use information contained in roll votes to assess their performance (see Bond and Fleisher 1990, chap. 3). Recognizing the limitations of roll call votes as a measure of legislative performance, for this analysis, votes supporting and opposing the president provide an appropriate parallel to runs scored and allowed.
5. The model and fuller discussion of findings are in Appendix A.
6. Using a 95% confidence interval, absolute values greater than 2.017 would be unusual; for our sample of 62 years, we would expect about 3.01 to appear unusual by random chance.
7. Miller (2007) discusses the mathematical theory underlying James' PE formula.
8. Hence the name "Pythagorean Expectation" because of its resemblance to the Pythagorean theorem:  $a^2 + b^2 = c^2$ .
9. Darowski uses WAE to construct a baseball manager Hall of Fame. By WAE, Mike Scioscia of the Angels is the best manager in the history of baseball. Bruce Bochy, Wilbert Robinson, Bobby Cox, and Filipe Alou round out the top five. <http://darowski.com/hall-of-wwar/expectancy/>.
10. Coauthor Teodoro believes that the 1984 Tigers' Jack Morris and Alan Trammell are Hall-of-Fame quality players, and that the Baseball Writers Association of America's (BWAA) failure to admit them to the Hall is an indictment of the BWAA's collective judgment.
11. There are others. Roll call votes, for example, don't involve the use of baseball bats, though Rep. Preston Brooks (D-SC) did beat Sen. Charles Sumner (R-MA) with a cane on the Senate floor in 1856 (US Senate n.d.).
12. As in the regression analysis, we exclude cloture votes in the Senate, as well as "consensus" roll calls on which the president won over 90 percent of the votes. Excluding virtual consensus votes from the calculation is consistent with the Freneau Efficiency Index (FEI), a popular performance metric for American college football, which is intended to measure "quality performances against good teams" (<http://www.footballoutsiders.com/stats/fei/>). The FEI calculation excludes "garbage time" plays—e.g., clock-killing kneel-downs or plays with third-string substitutes when the score is lopsided near the end of a game.
13. As a robustness check, we replicated the optimization procedure using the evolutionary stochastic optimization algorithm recommended by Yeniyay (2005). The values of  $\gamma$  were equal to the seventh decimal place.
14. Bootstrapped replications of the correlations between PE% and actual winning percentage clustered by Congress yielded confidence intervals of .912–.966 in the House and .752–.907 in the Senate.
15. The  $\gamma$  for NFL football and NBA basketball are 2.37 and 13.91, respectively, reflecting smaller average scoring margins in football and larger margins in basketball.
16. Appendix B presents results in presidential baseball cards. We thank J. Antonio Teodoro for useful suggestions in development of the presidential baseball cards.
17. Our choice to rank presidents by career WAE rather than WAE% rewards total legislative production during a presidency (WAE) over efficiency (WAE%). Measured by WAE%, Ford (not Reagan) is the top-ranked president and Clinton (not Eisenhower) finishes in last place. The difference between WAE and WAE% as gauges of greatness parallels another perennial debate among baseball fans: a player's career value versus his peak value. Cy Young is the iconic example of career value. Over his legendary 22-year career (1890–1911), Young collected every significant career record for pitching, many of which still stand more than a century after he retired. The archetypical example of peak value in baseball is

Sandy Koufax, who played 12 seasons (1955–1966), the first eight of which were unremarkable. But from 1963–1966 Koufax had arguably the most dominant four-year run by a pitcher and earned the nickname “The Left Arm of God” before his career was ended by injury. Historically, the baseball Hall of Fame voters have tended to favor career achievements, which naturally favors players who have long careers (<http://www.fangraphs.com/blogs/trammell-yount-and-the-value-of-career-length/>). We follow suit here, but we might have subtitled this paper “Why Gerald Ford is like Sandy Koufax and Ronald Reagan is like Cy Young.”

18. Darowski offers two WAE methodologies, one based on pure PE and another based on roster quality. Without comparable “roster quality” metrics for Congress, our rankings rely entirely on PE.

## REFERENCES

- Berra, Yogi. 2008. *You Can Observe a Lot by Watching: What I've Learned About Teamwork From the Yankees and Life*. New York: Wiley.
- Bond, Jon R. and Richard Fleisher. 1990. *The President in the Legislative Arena*. Chicago: University of Chicago Press.
- Cohen, Jeffrey E., Jon R. Bond, and Richard Fleisher. 2013. “Placing Presidential-Congressional Relations in Context: A Comparison of Barack Obama and His Predecessors.” *Polity* 45 (January): 105–26.
- Covington, Cary R., J. Mark Wrighton, and Rhonda Kinney. 1995. “A Presidency-Augmented Model of Presidential Success on House Roll Call Votes.” *American Journal of Political Science* 39 (November): 1001–24.
- Covington, Cary R. 1987a. “Mobilizing Congressional Support for the President: Insights from the 1960s.” *Legislative Studies Quarterly* 12 (February): 77–96.
- . 1987b. “Staying Private: Gaining Congressional Support for Unpublicized Presidential Preferences on Roll Call Votes.” *Journal of Politics* 49 (August): 737–55.
- . 1988a. “Building Presidential Coalitions Among Cross-pressured Members of Congress.” *Western Political Quarterly* 41 (March): 47–62.
- . 1988b. “Guess Who’s Coming to Dinner: The Distribution of White House Social Invitations and Their Effects on Congressional Support.” *American Politics Quarterly* 16 (July): 243–65.
- Edwards, George C. III. 2009. *The Strategic President: Persuasion and Opportunity in Presidential Leadership*. Princeton, NJ: Princeton University Press.
- Fett, Patrick J. 1994. “Presidential Legislative Priorities and Legislators’ Voting Decisions: An Exploratory Analysis.” *Journal of Politics* 56 (May): 502–12.
- Fleisher, Richard and Jon R. Bond. 1983. “Assessing Presidential Support in the House: Lessons from Reagan and Carter.” *Journal of Politics* 45 (August): 745–58.
- . 1992. “Assessing Presidential Support in the House II: Lessons from George Bush.” *American Journal of Political Science* 37 (May): 525–41.
- Fleisher, Richard, Jon R. Bond, and B. Dan Wood. 2008. “Which Presidents Are Uncommonly Successful in Congress?” In *Presidential Leadership: The Vortex of Power*, ed. Bert A. Rockman and Richard W. Waterman, 191–214. New York: Oxford University Press.
- Grabner, David. n.d. “The Sabermetric Manifesto.” *The Baseball Archive*, ed. Sean Lahman. <http://seanlahman.com/baseball-archive/sabermetrics/sabermetric-manifesto/>. Accessed September 21, 2016.
- James, Bill. 1982. *The Bill James Baseball Abstract, 1982*. New York: Ballantine Books.
- Kellerman, Barbara. 1984. *The Political Presidency: Practice of Leadership from Kennedy through Reagan*. New York: Oxford University Press.
- King, Anthony. 1983. “A Mile and Half Is a Long Way.” In *Both Ends of the Avenue: The Presidency, the Executive Branch, and Congress in the 1980s*, ed. Anthony King, 246–73. Washington, DC: American Enterprise Institute.
- Lockerbie, Brad and Stephen A. Borrelli. 1989. “Getting Inside the Beltway: Perceptions of Presidential Skill and Success in Congress.” *British Journal of Political Science* 19 (January): 97–106.
- Mantell, J. B. and Leon S. Lasdon. 1978. “A GRG Algorithm for Econometric Control Problems.” *Annals of Economic and Social Measurement* 6 (5): 581–97.
- Miller, Steven J. 2007. A Derivation of the Pythagorean Won-Loss Formula in Baseball. *Chance* 20 (1): 40–48.
- Neustadt, Richard E. 1960. *Presidential Power: The Politics of Leadership*. New York: Wiley.
- Peterson, Mark A. 1990. *Legislating Together: Eisenhower to Reagan*. Cambridge, MA: Harvard University Press.
- Rottinghaus, Brandon and Justin S. Vaughn. 2015. “Measuring Obama against the Great Presidents.” *FixGov: Making Government Work*. Washington, DC: Brookings Institution. February 13, 2015. <https://www.brookings.edu/2015/02/13/measuring-obama-against-the-great-presidents/> Accessed September 2, 2016.
- Sullivan, Terry. 1988. “Headcounts, Expectations, and Presidential Coalitions in Congress.” *American Journal of Political Science* 32 (August): 567–89.
- . 1990. “Explaining Why Presidents Count: Signaling and Information.” *Journal of Politics* 52 (August): 939–62.
- . 1991. “The Bank Account Presidency: A New Measure of Evidence on the Temporal Path of Presidential Influence.” *American Journal of Political Science* 35 (August): 686–723.
- US Senate n.d. “Senate Stories.” [http://www.senate.gov/artandhistory/history/minute/The\\_Canining\\_of\\_Senator\\_Charles\\_Sumner.htm](http://www.senate.gov/artandhistory/history/minute/The_Canining_of_Senator_Charles_Sumner.htm) Accessed 9-2-16.
- Yeniay, Ozgur. 2005. “A Comparative Study on Optimization Methods for the Constrained Nonlinear Programming Problems.” *Mathematical Problems in Engineering* 2: 165–173.