BRANCH RICKEY'S EQUATION FIFTY YEARS LATER

By

Ray C. Fair and Danielle Catambay

July 2005
Revised January 2007

COWLES FOUNDATION DISCUSSION PAPER NO. 1529

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Branch Rickey's Equation Fifty Years Later

Ray C. Fair[a] and Danielle Catambay[b*]

Revised January 2007

Abstract

This paper analyzes Branch Rickey's 1954 equation in a regression context. The results for 1934–1953 are consistent with Rickey's conclusions, and the equation holds up well when extended 51 years. Two of Rickey's main points were that on base percentage dominates batting average and that offense and defense are equally important, and these, along with the entire equation, are generally supported by the results. Rickey does seem to have been ahead of his time.

1 Introduction

Branch Rickey (1954) in a *Life* magazine article introduced an equation relating a baseball team's performance in a season to various measures of offense and defense. One of his findings was that on base percentage dominates batting average in the measure of offense, which, as Schwarz (2004a) notes, was way ahead of its time. Rickey's analysis is quite interesting. It is probably largely due to Allan Roth, whom he mentions in the article. Rickey and Roth were not mathematical

---

[*a] Cowles Foundation and International Center for Finance, Yale University, New Haven, CT 06520-8281. Voice: 203-432-3715; Fax: 203-432-6167; email: ray.fair@yale.edu; website: http://fairmodel.econ.yale.edu. [b] Yale University; email: danielle.catambay@yale.edu. This paper was stimulated by Alan Schwarz's (2004a) article in the *New York Times*. We are indebted to Sharon Oster for helpful comments.

statisticians, and they took their figures "to mathematicians at a famous research institute" (p. 79) (alas, Princeton, not Yale[1]). They got their results back in six weeks, "which constituted a framework around which to build a formula" (p. 79). Rickey does not discuss in a mathematically rigorous way the derivation of his formula, but there is enough discussion of technique in the article to see roughly what he did.

In this paper Rickey's equation is examined using a more formal statistical technique, regression analysis, which is often used in the social sciences. The equation is first examined using Rickey's own period, 1934–1953, and then it is extended to the present to see how it does in the modern era. It will be seen that the results for 1934–1953 support Rickey's conclusions and that the equation holds up well when extended 51 years through 2004. Although Rickey's equation was largely ignored at the time, the results in this paper suggest that perhaps it should not have been. The equation is presented in Section 2 and then analyzed in Section 3.

2 The Equation

Rickey said he used the last 20 years worth of data to build his formula; we will assume that 1934–1953 were the 20 years in question. The data are yearly and by team.[2] In this period there were 16 teams, 8 per league, and so the number of

[1] Schwarz (2004b), p. 58.

[2] The data for this paper were downloaded from the website: http://baseball1.com. Team data were available for all but hit by pitch, HP, opponents hit by pitch, HB^*, and opponents at bats, AB^*. These three variables were constructed from individual player data, which were available. HP was constructed as the sum of the number of times each player on the team was hit by a pitch.

observations we can use is 320. Rickey used as the measure of team performance the number of games behind the league leader for the season, denoted G. He was also interested in a team's average runs per game in a season relative to the average runs per game of the team's opponents. Rickey first noted that this variable and G are highly positively correlated. This, of course, is not surprising. The more runs a team scores relative to its opponents, the more games it is likely to win. Rickey's aim was then to see if he could find measures of offense that were highly correlated with a team's average runs per game and measures of defense that were highly correlated with the average runs per game of the team's opponents. Such measures would then be highly correlated with G and would give one an idea of the *kinds* of offense and defense that are most effective. In the end (after getting back the results from the mathematical experts) he came up with three measures of offense and four measures of defense.

The first measure of offense is on base percentage:

$$onbase = \frac{H + BB + HP}{AB + BB + HP} \tag{1}$$

where H is hits, BB is bases on balls, HP is hit by pitch, and AB is at bats.[3] These variables are all a team's totals for the season. The second is a measure of extra base power:

$$power = 0.75\frac{TB - H}{AB} \tag{2}$$

where TB is total bases (4 times home runs plus 3 times triples plus 2 times doubles

---

HB^* was constructed as the sum of the number of times each pitcher on the team hit a batsman. AB^* was constructed as the sum of the number of batsmen faced by each pitcher on the team. A data base from 1921 through 2004 was created.

[3]The modern definition of on base percentage adds sacrifice flies to the denominator. In this paper we use only Rickey's definitions.

plus singles). Rickey said that $(TB-H)/AB$ had a lower correlation with a team's average runs per game than did the other two measures (the third one discussed next), and he adjusted for this by multiplying it by 0.75. We return to this in the next section. The third measure is what Rickey calls "clutch:"

$$clutch = \frac{R}{H + BB + HP} \tag{3}$$

where R is runs scored. This variable is the percent of players on base who score.[4] The total offense measure is then the sum of the three:

$$offense = onbase + power + clutch \tag{4}$$

Defense was calculated by Rickey using four measures. These are measures that are meant to be highly correlated with the average runs per game of a team's opponents. The first measure is opponents' batting average:

$$oppba = \frac{H^*}{AB^*} \tag{5}$$

where H^* is hits by opponents and AB^* is at bats by opponents. The second measure is the percentage of opponents who get on base because of walks or hit batsmen:

$$oppbb = \frac{BB^* + HB^*}{AB^* + BB^* + HB^*} \tag{6}$$

where BB^* is bases on balls by opponents and HB^* is the number of opponents hit by a pitch. The third measure is a "clutch" measure for pitching: the percentage of base runners scoring earned runs for the opponents

$$opper = \frac{ER^*}{H^* + BB^* + HB^*} \tag{7}$$

[4]Rickey is excluding here players who get on base because of an error or interference.

where ER^* is earned runs scored by the opponents. Finally, the fourth measure is strikeout percentage:

$$oppso = -0.125 \frac{SO^*}{AB^* + BB^* + HB^*} \tag{8}$$

where SO^* is the number of opponent strike outs. Rickey did not find strike outs to be of "equal importance" to the others, and he weighted the strikeout percentage by only 0.125. We also return to this in the next section. Note that there is a minus sign in front of 0.125: the more strikeouts, the worse are the opponents. The total defense measure is then the sum of the four:

$$defense = oppba + oppbb + opper + oppso \tag{9}$$

Rickey's final equation is then:

$$G = offense - defense \tag{10}$$

Rickey also adds fielding, denoted F, to this equation. However, he has no measure of F, and F plays no role in the article. We will thus ignore F in this paper.[5]

The formula given in (10) is, of course, not literally an equation explaining G. Rickey was dealing with correlations, and it is not the case that the coefficient of $offense$ should be one and that of $defense$ minus one. Among other things, the signs are wrong. $offense$ should have a negative effect on G and $defense$ a positive effect, since G is the number of games behind. Rather, Rickey's equation should be looked upon as a guide to what he thought was important in helping a

[5]Rickey states that "There is nothing on earth anybody can do with fielding" and "Fielding then cannot be measured" (p. 81). However, he then goes on to say "But application of the formula to 20 years of statistics shows fielding to be worth only about one half as much as pitching or about 15%" (p. 81). How he knows this if fielding cannot be measured is unclear.

baseball team win games. In the next section we put Rickey's baseball expertise to a more rigorous statistical test.

3 Regression Analysis

From a formal statistical perspective, Rickey's formula offers a number of predictions. First, in explaining games behind the leader, G, the offense and defense measures that matter most are $onbase$, $power$, $clutch$, $oppba$, $oppbb$, $opper$, and $oppso$. A stronger prediction is *how* these measures should matter. Rickey's explanation is that the three offense measures should matter equally, as should the four defense measures. We can test these predictions using the following equation:

$$\begin{aligned} G_{it} &= \gamma + \alpha_1 onbase_{it} + \alpha_2 power_{it} + \alpha_3 clutch_{it} + \beta_1 oppba_{it} + \beta_2 oppbb_{it} \\ &+ \beta_3 opper_{it} + \beta_4 oppso_{it} + u_{it}, \quad i = 1, \ldots, 16, \quad t = 1934, \ldots, 1953 \end{aligned} \tag{11}$$

where the it subscript has been added to the variables to denote that each is for team i and year t. If Rickey's view is right, then the α's should equal each other and the β's should equal each other, and this can be tested.

The results of estimating equation (11) by ordinary least squares (regression analysis) are presented in Table 1. Two sets of estimates are presented: one unrestricted and one with the α and β restrictions, as predicted by Rickey. Presented in brackets below the variables are the partial correlation coefficients. A partial correlation coefficient measures the correlation of the variable with G after the effects of all the other variables have been taken into account. t-statistics are also presented in the table. A variable is considered to be statistically significant if its

Table 1
Coefficient Estimates
Sample period: 1934–1953

$$(11)\ \ G = \gamma + \alpha_1 onbase + \alpha_2 power + \alpha_3 clutch + \beta_1 oppba + \beta_2 oppbb + \beta_3 opper + \beta_4 oppso + u$$

or

$$(11)'\ \ G = \gamma + \alpha_1 offense + \beta_1 defense + u$$

	Estimate of									
	γ	α_1	α_2	α_3	β_1	β_2	β_3	β_4	R^2	# obs.
(11)	22.6	−302.2	−30.9	−155.0	317.2	362.1	193.0	603.6	.823	320
	(1.60)	(−9.01)	(−0.96)	(−6.04)	(5.82)	(8.33)	(7.92)	(1.98)		
		[-.455]	[-.054]	[-.324]	[.313]	[.427]	[.409]	[.111]		
(11)′	−8.8	−150.1			249.5				.801	320
	(−1.28)	(−26.03)			(27.59)					
		[-.825]			[.840]					

• t-statistics in parentheses, partial correlation coefficients in brackets.
• Standardized coefficients for equation $(11)'$ for $offense$ and $defense$ are -0.657 and 0.696, respectively.
• When batting average, H/AB, is added to equation (11), the t-statistic is -0.05.

t-statistic is greater than about 2.0 in absolute value. In the following discussion "p-values" are sometimes mentioned. p-values lie between 0 and 1. The larger the p-value for a test the more confidence one can have that the hypothesis being tested is true. A hypothesis is generally considered rejected if the p-value is 0.05 or less.

Consider the unrestricted results in Table 1 first—equation (11). The partial correlation coefficients are similar for all but $power$ and $oppso$, ranging in absolute value between 0.313 and .455. They are much smaller for $power$ and $oppso$, which is what Rickey said he found and which led him to weight them less in the equation. Looking at the t-statistics, all the variables are statistically significant except for $power$.

Comparing now the restricted results with the unrestricted ones—equation

(11) versus (11)′—for the restricted equation the R^2, a measure of the overall fit of the equation, fell from 0.823 to 0.801. The two variables, $offense$ and $defense$, are highly significant. The coefficient estimate of $offense$ is not close to that of $defense$ in absolute value (-150.1 versus 249.5), but the standardized coefficients are: -0.657 and 0.696. Standardized coefficients are adjusted for the variation in the variables, which in the present context is useful to do. These similar standardized coefficients (in absolute value) say that a typical change in $offense$ has a similar effect on G as a typical change in $defense$ (with the sign reversed).

One of the more interesting results for the restricted equation in Table 1 is that the partial correlation coefficients are close in absolute value: -0.825 and 0.840. This closeness is consistent with Rickey's discussion of offense versus defense. One of his main points was that offense and defense were equally important, much to his and other people's surprise.[6] It is not clear in the article how Rickey arrived at this conclusion, but perhaps it was from observing (through the mathematicians) the closeness of these correlations.

An F test can be used to test if the decrease in fit in moving from the unrestricted to the restricted equation in Table 1 is statistically significant. The hypothesis that is tested using the F test is that α's are all equal to each other and the β's are all equal to each other. This hypothesis is rejected.[7] It is, however, not clear whether this rejection should count against Rickey's equation, because it is not clear why Rickey added the three offense variables and the four defense variables together in the first

[6]When George Sisler saw the figures "his reaction was one of bewilderment. 'I still don't believe it,' he said. 'But there it is"' (p. 83).

[7]The F value was 7.45, which with 5, 312 degrees of freedom has a p-value that is zero to over three decimal places.

place. He was looking for variables that were highly correlated with a team's average runs per game and the average runs per game of the team's opponents, not necessarily variables with similar coefficients in an equation like (11). He did weight $power$ by 0.75 because of what he said was its lower correlation. The unrestricted estimates in Table 1 show that this weight was not low enough if one were looking for a coefficient estimate for $power$ close to those for $onbase$ and $clutch$. On the other hand, the weight he used for $oppso$, 0.125, was too low if he were looking for similar coefficient estimates for the defense variables because the coefficient estimate for $oppso$ is noticeably larger than the others.

Although both the unrestricted and restricted estimates are presented in Table 1 (and in Table 2 below), we will take the regression version of Rickey's equation to be the unrestricted equation, namely equation (11). In other words, we will give Rickey the benefit of the doubt and assume that he was looking for significant variables and not necessarily variables with the same coefficient in an equation like (11).

Rickey was right in that on base percentage is a better measure than batting average for offense. When batting average, H/AB, is added to the unrestricted equation, it has a t-statistic of only -0.05. $onbase$ completely dominates.

It is interesting that for defense Rickey did not use on base percentage. He used opponents' batting average, $oppba$, and the percentage of opponents who get on base because of walks or hit batsmen, $oppbb$. If on base percentage were used, the variable would be:

$$opponbase = \frac{H^* + BB^* + HB^*}{AB^* + BB^* + HB^*} \tag{12}$$

and $opponbase$ would replace $oppba$ and $oppbb$ in equation (11). Testing for $opponbase$ versus $oppba$ and $oppbb$ is what is called a "nonnested test" in statistics. One test that can be used is the Davidson-MacKinnon (1981) test. This test takes the fitted values from equation (11) and adds them as an explanatory variable to the equation with $opponbase$ included and $oppba$ and $oppbb$ excluded. When this was done for the 1934–1953 sample period the t-statistic for the fitted values was 2.30, which has a p-value of 0.022. The fitted values are thus significant. Conversely, when the fitted values from the equation with $opponbase$ included and $oppba$ and $oppbb$ excluded were added to the equation with $oppba$ and $oppbb$ included and $opponbase$ excluded, the t-statistic for the fitted values was -0.61, which has a p-value of 0.544. These fitted values are thus not significant. Because the first fitted values are significant and the second not, this test rejects $opponbase$ in favor of $oppba$ and $oppbb$. Once again Rickey seems to have made the right choice.

Overall, the results in Table 1 seem supportive of Rickey's analysis. The next step is to see how the equation fares over time. Table 2 presents results of estimating the equation through 2004. For these results the left hand side variable, the variable to be explained, was changed from games behind to games behind as a percent of the number of games played in that season by the league leader, denoted GP. This adjusts for the 1961 increase in the number of games played in a season from 154 to 162. Also, in computing games behind, divisions within a league (when they exist) were combined, making just two leagues. Three sets of estimates are presented in Table 2. The first for the entire 1934–2004 period; the second for Rickey's period, 1934–1953; and the third for the period beyond Rickey's, 1954–2004.

The same conclusions hold for the entire period as hold for Rickey's period,

Table 2
Coefficient Estimates with GP as Dependent Variable
Sample periods: 1934–2004, 1934–1953, 1954–2004

$$(13)\ \ GP == \gamma + \alpha_1 onbase + \alpha_2 power + \alpha_3 clutch + \beta_1 oppba + \beta_2 oppbb + \beta_3 opper + \beta_4 oppso + u$$

or

$$(13)'\ \ GP = \gamma + \alpha_1 offense + \beta_1 defense + u$$

	Estimate of									
	γ	α_1	α_2	α_3	β_1	β_2	β_3	β_4	R^2	# obs.
(13) 1934-2004	0.070	-1.80	-0.12	-1.13	2.18	2.85	1.13	2.25	.767	1548
	(1.88)	(-19.45)	(-1.36)	(-14.83)	(14.53)	(22.49)	(16.52)	(5.13)		
		[-.440]	[-.035]	[-.354]	[.347]	[.497]	[.388]	[.130]		
$(13)'$ 1934-2004	-0.094	-0.93			1.62				.728	1548
	(-4.96)	(-50.33)			(55.54)					
		[-.788]			[.816]					
(13) 1934-1953	0.137	-1.96	-0.21	-1.00	2.08	2.36	1.26	3.83	.822	320
	(1.49)	(-8.98)	(-1.00)	(-5.97)	(5.86)	(8.32)	(7.95)	(1.93)		
(13) 1954-2004	0.090	-1.84	0.03	-1.25	2.08	2.95	1.13	0.79	.749	1228
	(2.14)	(-17.65)	(0.27)	(-14.32)	(12.55)	(20.86)	(15.00)	(1.46)		

- GP for a team and year is G divided by the number of games played by the league leader.
- t-statistics in parentheses, partial correlation coefficients in brackets.
- Standardized coefficients for equation $(13)'$ for $offense$ and $defense$ are -0.718 and 0.792, respectively.
- When batting average, H/AB, is added to equation (13) for the 1934-2004 sample period, the t-statistic is -0.57.

namely 1) that $power$ is not significant, 2) that all but $power$ and $oppso$ have similar partial correlation coefficients, and 3) that when only $offense$ and $defense$ are explanatory variables they have similar partial correlation coefficients. Also, when batting average, H_{it}/AB_{it}, is added to the unrestricted equation, it has a t-statistic of only -0.57. Again, $onbase$ completely dominates. The coefficient estimates for the two sub samples are fairly close except for $oppso$ and perhaps $oppbb$. The hypothesis that the coefficients in the two sub samples except for the constant term are equal can be tested using an F test. This test yielded an F value of 2.45, which with 7, 1532 degrees of freedom has a p-value of 0.017. This p-value is less than 0.05, and so by conventional standards the hypothesis is rejected. The hypothesis of equality is not rejected if the cutoff is taken to be 0.01, which in practice it sometimes is. So the decision in this case is close. Overall, the results in Table 2 show that Rickey's equation holds up quite well when extended 51 years.[8]

4 Conclusion

Although Branch Rickey's *Life* article is full of hyperbole and the discussion of how he arrived at his conclusions is somewhat murky, the statistical results in this paper generally support his choices. The variables that he ended up choosing except for $power$ are statistically significant when tested in a regression context, and the correlation framework has not changed much over time. Rickey's conclusion that batting average is dominated by on base percentage is confirmed, and his conclusion that offense and defense are equally important is confirmed in that the $offense$ and $defense$ variables have similar partial correlation coefficients in absolute value. The subtitle to the *Life* article is "'The Brain' of the game unveils formula that statistically disproves cherished myths and demonstrates what really wins." It looks like he did.

[8] A few other tests that were performed are the following. When various stability tests like the one reported above were performed, the F values tended to be fairly low, but the p-values were sometimes less than 0.01. For example, when the sample is extended back to 1921 and the hypothesis that the coefficients in the three sub samples, 1921–1933, 1934–1953, and 1954–2004, are equal (except for the constant term), the F value is 2.42, which with 14, 1732 degrees of freedom has a p-value of 0.002. For the sample period 1934–2004 the hypothesis that the coefficients for the American league teams are the same as those for the National league teams was tested, and the F value was 0.98, which with 7, 1532 degrees of freedom has a p-value of 0.441. The hypothesis of stability between the American and National leagues is thus not rejected.

References

[1] Davidson, R. and J. G. MacKinnon, 1981, "Several Tests of Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49, 781-793.

[2] Rickey, Branch, 1954, "Goodby to Some Old Baseball Ideas," *Life*, August 2, 1954, 79-89.

[3] Schwarz, Alan, 2004a, "Looking Beyond Batting Average," *The New York Times*, August 1, 2004.

[4] Schwarz, Alan, 2004b, *The Numbers Game*. New York: St. Martin's Press.