SUBSCRIPTIONS

NINE: A Journal of Baseball History and Culture (ISSN 1188-9330) is published semiannually at $50 for individuals, $85 for institutions, and $27 for single issues by the University of Nebraska Press. For subscriptions outside the United States, please add $20 for postage. Canadian subscribers, add appropriate GST or HST. Residents of Nebraska, please add the appropriate Nebraska sales tax. Payment must accompany order. Make checks payable to the University of Nebraska Press and mail to:
PO Box 84555
Lincoln, NE 68501-4555
Telephone 800-755-1105 (United States and Canada)
402-472-3581 (other countries)
www.nebraskapress.unl.edu

SUBMISSIONS

NINE seeks to promote the study of all historical aspects of baseball and centers on the cultural implications of the game wherever in the world baseball is played. The journal reflects an eclectic approach and does not foster a particular ideological bias.

Manuscripts must be prepared according to the Chicago Manual of Style, 15th edition, and should contain endnotes rather than footnotes. If an article prepared in another style is accepted, the author is fully responsible for changing the style to CMS. Manuscripts prepared in American Psychological Association (APA) style are not acceptable. Submissions can be e-mailed to tstrecker@bsu.edu.

Simultaneous submissions are not accepted. Authors will be asked to sign a declaration that the submitted work is not being considered for publication elsewhere. No major revisions will be allowed after acceptance. The editor will attempt to have all manuscripts reviewed by two umpires and will respond to the authors within three months of receiving the manuscript.

Individuals whose works are accepted for publication must supply them in both paper and electronic form (Microsoft Word or RTF). Articles should be between 2,000 and 5,000 words in length.

All figures and tables must be submitted in camera-ready format and will be printed as is. Please include a short (no more than three sentences) personal bio.

Address submissions to:
Trey Strecke
Ball State University
Department of English
Muncie, IN 47306-0460

Book reviews are by invitation only.
The views expressed in NINE are those of the contributors and not necessarily those of the editor and publishers.

©2008 by the University of Nebraska Press
All rights reserved
Manufactured in the United States of America


"Baseball's Limitless Dramatic Possibilities: Keynote Speech to the Fourteenth Annual NINE Spring Training Conference, March 17, 2007" © 2008 by Robert Creamer

"Only Peggy Lee Could Sing of My Mets Misery" © 2008 by Mary Herbert

With the exception of brief passages, no part, in any form or in any means, may be reproduced without permission in writing from the University of Nebraska Press, 1111 Lincoln Mall, Lincoln, NE 68588-0630.

Articles appearing in this journal are abstracted and indexed in Academic Search Premier, America: History & Life, Current Abstracts, SocINDEX, SPORTDiscus, and TOC Premier.

NINE is available online through Project MUSE at http://muse.jhu.edu.
FOUNDING EDITOR

Bill Kirwin,
The University of Calgary in Edmonton

EDITOR

Trey Strecker,
Department of English
Ball State University
Muncie, IN 47306-0460

EDITORIAL BOARD

Robert K. Barney, University of Western Ontario, London, Ontario
Robert Bellamy, Duquesne University, Pittsburgh, Pennsylvania
Adrian Burgos Jr., University of Illinois, Urbana, Illinois
Peter Carino, Indiana State University, Terre Haute, Indiana
Richard C. Crepeau, University of Central Florida, Orlando, Florida
Gerald Early, Washington University, St. Louis, Missouri
Larry Gerlach, University of Utah, Salt Lake City, Utah
Steven P. Gietschier, St. Louis, Missouri
George Gmelch, University of San Francisco, San Francisco, California; Union College, Schenectady, New York
Chris Lamb, College of Charleston, Charleston, South Carolina
David Mills, University of Alberta, Edmonton, Alberta
Roberta Newman, New York University, New York City, New York
James Odenkirk, Arizona State University, Tempe, Arizona
Joel Nathan Rosen, Moravian College, Bethlehem, Pennsylvania
Jules Tygiel, San Francisco State University, San Francisco, California
Clearing the Bases
The Greatest Baseball Debates of the Last Century
ALLEN BARRA
Foreword by Bob Costas
$16.95 paper

Field of Schemes
How the Great Stadium Swindle Turns Public Money into Private Profit, Revised and Expanded Edition
NEIL DEMAUVE
AND JOANNA CAGAN
$19.95 paper

A BISON ORIGINAL
Venezuelan Bust, Baseball Boom
Andrés Reiner and Scouting on the New Frontier
MILTON H. JAMAAL
$19.95 paper

Ed Barrow
The Bulldog Who Built the Yankees' First Dynasty
DANIEL R. LEVITT
$29.95 cloth

Chief Bender's Burden
The Silent Struggle of a Baseball Star
TOM SWIFT
$24.95 cloth

A BISON ORIGINAL
Center Field Shot
A History of Baseball on Television
JAMES R. WALKER AND ROBERT V. BELLAMY JR.
$24.95 paper

*For complete descriptions and to read excerpts, visit us online!
Contents

NINE FALL 2008, VOL. 17, NO. 1

ix Diamond Quotes

ARTICLES

1 The Saints on Sunday: Ballparks and Sabbatarianism in St. Paul
   KRISTIN M. ANDERSON AND CHRISTOPHER W. KIMBALL

16 The Optimist: Baseball Themes in Bert J. Griswold’s
   Fort Wayne Editorial Cartoons
   CHRISTOPHER BAAS

37 Rumors and Facts: William Clarence Matthews’s
   1905 Challenge to Major League Baseball’s Color Barrier
   KARL LINDHOLM

54 Baseball and Dissent: The Vietnam Experience
   RON BRILEY

70 Baseball’s Great Hitting Barrage of the 1990s
   (and Beyond) Reexamined
   BENJAMIN G. RADER AND KENNETH J. WINKLE

TRIPLE PLAY

97 Baseball’s Limitless Dramatic Possibilities:
   Keynote Speech to the Fourteenth Annual NINE
   Spring Training Conference, March 17, 2007
   ROBERT CREAMER
Branch Rickey's Equation Fifty Years Later
RAY C. FAIR AND DANIELLE CATAMBAY

From the Dugout to the Classroom:
Why Good Baseball Coaches Have Much to Offer Good Professors
DARIN H. VAN TASSELL

POETRY

Only Peggy Lee Could Sing of My Mets Misery
MARY HERBERT

Goose Feathers
STEVE KASH

BOOK REVIEWS

Branch Rickey: Baseball's Ferocious Gentleman
BILL MARSHALL

The Curt Flood Story: The Man Behind the Myth
RICHARD CREPEAU

Sal Maglie: Baseball's Demon Barber
MARK ARMOUR

Out of the Park: Memoir of a Minor League Baseball All-Star
STEVE GIETSCHIER

Level Playing Fields: How the Groundskeeping
Murphy Brothers Shaped Baseball
CHRIS LAMB

Journeymen
ERIC B. SALO

Scoring From Second: Writers on Baseball
SCOTT D. PETERSON

The Arrival of the American League: Ban Johnson and
the 1901 Challenge to National League Monopoly
HARRY JEBSEN JR.

When Chicago Ruled Baseball: The Cubs-White Sox World Series of 1906
DAVID A. GOSS
149 Connie Mack and the Early Years of Baseball
   JAMES E. MARTIN

150 The Gashouse Gang: How Dizzy Dean, Leo Durocher, Branch Rickey, Pepper Martin, and Their Colorful, Come-from-Behind Ball Club Won the World Series—and America's Heart—During the Great Depression
   DOUGLAS K. LEHMAN

151 The Soul of Baseball: A Road Trip Through Buck O'Neil's America
   HAROLD V. HIGHAM

152 Playing in Isolation: A History of Baseball in Taiwan
   ROBERTA NEWMAN

155 Growing the Game: The Globalization of Major League Baseball
   KEVIN G. QUINN

FILM REVIEW

158 The Brooklyn Dodgers: The Ghosts of Flatbush
   FRANK ARDOLINO

ORAL HISTORY

164 The Alomars: On Baseball, Parenting, and Being Puerto Rican
   FRANK OTTO

170 George Sheldon: Marketing the Aluminum Bat
   GEORGE GMELCH

176 Contributors
Wally Yonamine  
The Man Who Changed Japanese Baseball  
ROBERT K. FITTS  
Foreword by  
Senator Daniel K. Inouye  
$26.95 cloth

Shoeless Joe and Ragtime Baseball  
HARVEY FROMMER  
With a new introduction  
by the author  
$16.95 paper

The Warsaw Sparks  
A Memoir  
GARY GILDNER  
Introduction by Albert E. Stone  
$15.95 paper

A BISON ORIGINAL  
Inside the Ropes  
Sportswriters Get Their Game On  
EDITED AND WITH AN  
INTRODUCTION BY  
ZACHARY MICHAEL JACK  
$24.95 paper

The Postwar Yankees  
Baseball's Golden Age Revisited  
DAVID G. SURDAM  
$45.00 cloth

Three Finger  
The Mordecai Brown Story  
CINDY THOMSON  
AND SCOTT BROWN  
Foreword by Ferguson Jenkins  
$17.95 paper

*For complete descriptions and to read excerpts, visit us online!

University of Nebraska Press  
800.755.1105 | www.nebraskapress.unl.edu | publishers of Bison Books
Branch Rickey’s Equation
Fifty Years Later

RAY C. FAIR AND DANIELLE CATAMBAY

INTRODUCTION

In a 1954 Life magazine article, Branch Rickey 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 notes, was way ahead of its time.¹ 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 statisticians, and they took their figures to “mathematicians at a famous research institute” (alas, Princeton, not Yale).² They got their results back in six weeks, “which constituted a framework around which to build a formula.”³ 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 data from Rickey’s own period, 1934 to 1953, and then it is extended to the present to see how it does with data from the modern era. It will be seen that the results from 1934 to 1953 support Rickey’s conclusions and that the equation holds up well when extended fifty-one 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

Rickey said he used the last twenty years worth of data to build his formula; we will assume that 1934 to 1953 were the twenty years in question. The data are yearly and by team.⁴ In this period, there were sixteen teams, eight per
league, so the number of 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 oppo-
nents. Rickey first noted that this variable and G have a strong positive cor-
relation. This 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 get-
ting 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

\[
onbase = \frac{H+BB+HP}{AB+BB+HP}\]

where H is hits, BB is bases on balls, HP is hit by pitch, and AB is at bats.\(^5\)
These variables are all a team’s totals for the season. The second equation is a
measure of extra base power

\[
\text{power} = 0.75 \frac{TB-H}{AB}\]

where TB is total bases (calculated by multiplying the number of home runs
by 4, the number of triples by 3, the number of doubles by 2, and adding
those figures to the number of 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 (equations 1 and 3), and he adjusted for this by multiplying it by
0.75, which is something we will return to later on. The third measure is what
Rickey calls “clutch”

\[
\text{clutch} = \frac{R}{H+BB+HP}\]

where R is runs scored. This variable is the percent of players on base who
score.\(^6\) The total offense measure is then the sum of these first three formulas:

\[
\text{offense} = \text{onbase} + \text{power} + \text{clutch}\]
Rickey used four measures to calculate defense. 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

\[ oppba = \frac{H^*}{AB^*} \]

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

\[ oppbb = \frac{BB^* + HB^*}{AB^* + BB^* + HB^*} \]

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

\[ opper = \frac{ER^*}{H^* + BB^* + HB^*} \]

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

\[ oppso = -0.125 \frac{SO^*}{AB^* + BB^* + HB^*} \]

where \( SO^* \) is the number of opponent strikeouts. Rickey did not find strikeouts to be of “equal importance” to the others, and he weighted the strikeout percentage by only 0.125—another figure we will soon return to. 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 these four defensive equations:

\[ defense = oppba + oppbb + opper + oppso \]

Rickey’s final equation is then:

\[ G = offense - defense \]
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.\(^7\)

The formula given in Rickey's final equation 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 1 and that of defense -1. 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 baseball team win games. We will now put Rickey's baseball expertise to a more rigorous statistical test.

REREGRESSION 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 offensive measures should matter equally, as should the four defensive ones. We can test these predictions using the following equation

\[
G_{it} = g + a_{1}onbase_{it} + a_{2}power_{it} + a_{3}clutch_{it} + b_{1}oppba_{it} + b_{2}oppbb_{it} + b_{3}opper_{it} + b_{4}oppso_{it} + u_{it}, \quad i = 1, \ldots, 16, \quad t = 1934, \ldots, 1953
\]

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 \( a \)s should equal each other and the \( \beta \)'s should equal each other, which 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 \( \alpha \) and \( \beta \) 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. Also presented in the table are t-statistics. A variable is considered to be statistically significant if its t-statistic is greater than about 2.0 in absolute value. In the following discussion "p-values" are sometimes mentioned. A p-value lies 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.
TABLE 1. Coefficient estimates. Sample period: 1934–1953

\[ G = \gamma + a_1 \text{onbase} + a_2 \text{power} + a_3 \text{clutch} + \beta_1 \text{oppba} + \beta_2 \text{oppbb} + \beta_3 \text{opper} + \beta_4 \text{oppso} + u \]

or

\[ G = \gamma + a_1 \text{offense} + \beta_1 \text{defense} + u \]

| \( \gamma \) | \( a_1 \) | \( a_2 \) | \( a_3 \) | \( \beta_1 \) | \( \beta_2 \) | \( \beta_3 \) | \( \beta_4 \) | \( R^2 \) | #obs.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>22.6</td>
<td>-302.2</td>
<td>-30.9</td>
<td>-155.0</td>
<td>317.20</td>
<td>362.1</td>
<td>193.0</td>
<td>603.6</td>
<td>.823</td>
</tr>
<tr>
<td>(1.60)</td>
<td>(-9.01)</td>
<td>(-0.96)</td>
<td>(-6.04)</td>
<td>(5.82)</td>
<td>(8.33)</td>
<td>(7.92)</td>
<td>(1.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11'</td>
<td>-8.8</td>
<td>-150.1</td>
<td></td>
<td>249.5</td>
<td></td>
<td></td>
<td></td>
<td>.801</td>
<td>320</td>
</tr>
<tr>
<td>(-1.28)</td>
<td>(-26.03)</td>
<td></td>
<td></td>
<td>(27.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-.825]</td>
<td></td>
<td></td>
<td></td>
<td>[.840]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: All t-statistics are enclosed in parentheses; partial correlation coefficients are enclosed in brackets. The standardized coefficients in 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.

Consider the unrestricted results for equation 11 in table 1 first. The partial correlation coefficients are similar for all but power and oppso, which are much smaller, ranging in absolute value between 0.313 and 0.455. This is what Rickey said he found and what led him to weigh 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 equation 11' in table 1—the \( R^2 \) figure, a measure of the overall fit of the equation, fell from 0.823 to 0.801 in the restricted equation. Thus 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. It is not clear in the article how Rickey arrived at this conclusion, but perhaps it was from observing (by way of 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 tested using the F-test is that the $a$s are all equal to each other and the $b$s are all equal to each other. This hypothesis is rejected.\(^9\) 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 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 as in equation 11. He did weigh 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 was 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 again 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, as in equation 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 averages, oppba, and the percentage of opponents who get on base because of walks or hit batters, oppbb. If on-base percentage were used, the variable would be

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

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 test.\(^{10}\) 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 sample period between 1934 and 1953, the t-statistic for the fitted values was 2.30, which has a p-value of 0.022. The
TABLE 2. Coefficient estimates with GP as dependent variable

\[ GP = \gamma + \alpha_{onbase} + \alpha_{power} + \alpha_{clutch} + \beta_{oppba} + \beta_{oppbb} + \beta_{opper} + \beta_{oppso} + u \]

or \[ 13' \text{ GP} = \gamma + \alpha_{offense} + \beta_{defense} + u \]

<table>
<thead>
<tr>
<th>ESTIMATE OF</th>
<th>( \gamma )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
<th>( \alpha_3 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( R^2 )</th>
<th>#obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>0.070</td>
<td>-1.80</td>
<td>-0.12</td>
<td>-1.13</td>
<td>2.18</td>
<td>2.85</td>
<td>1.13</td>
<td>2.25</td>
<td>.767</td>
<td>1548</td>
</tr>
<tr>
<td>1934–2004</td>
<td>(1.88)</td>
<td>(-19.45)</td>
<td>(-1.36)</td>
<td>(-14.83)</td>
<td>(14.53)</td>
<td>(22.49)</td>
<td>(16.52)</td>
<td>(5.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13'</td>
<td>-0.094</td>
<td>-0.93</td>
<td></td>
<td>1.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.728</td>
<td>1548</td>
</tr>
<tr>
<td>1934–2004</td>
<td>(-4.96)</td>
<td>(-50.33)</td>
<td></td>
<td>(55.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-.788]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.137</td>
<td>-1.96</td>
<td>-0.21</td>
<td>-1.00</td>
<td>2.08</td>
<td>2.36</td>
<td>1.26</td>
<td>3.83</td>
<td>.822</td>
<td>320</td>
</tr>
<tr>
<td>1934–1953</td>
<td>(1.49)</td>
<td>(-8.98)</td>
<td>(-1.00)</td>
<td>(-5.97)</td>
<td>(5.86)</td>
<td>(8.32)</td>
<td>(7.95)</td>
<td>(1.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.090</td>
<td>-1.84</td>
<td>0.03</td>
<td>-1.25</td>
<td>2.08</td>
<td>2.95</td>
<td>1.13</td>
<td>0.79</td>
<td>.749</td>
<td>1228</td>
</tr>
<tr>
<td>1954–2004</td>
<td>(2.14)</td>
<td>(-17.65)</td>
<td>(0.27)</td>
<td>(-14.32)</td>
<td>(12.55)</td>
<td>(20.86)</td>
<td>(15.00)</td>
<td>(1.46)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The GP for a team and year is \( G \) divided by the number of games played by the league leader. All t-statistics are enclosed in parentheses; partial correlation coefficients are enclosed in brackets. The standardized coefficients in 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.

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, namely (1) that power is not significant, (2) all but power and oppso have similar partial correlation coefficients, and (3) when only offense and defense are explanatory variables they have similar partial correlation coefficients. Also, when batting average, $H_u/AB_u$, 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 to the present.11

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.

NOTES


4. The data for this paper were downloaded from the Web site 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. 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 database from 1921 through 2004 was created.

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

6. Rickey excludes players who get on base because of an error or interference.

7. Rickey states that “There is nothing on earth anybody can do with fielding” and “Fielding then cannot be measured.” However, he 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%” (Branch Rickey, “Goodbye to Some Old Baseball Ideas,” Life, August 2, 1954, 81). How he knows this if fielding cannot be measured is unclear.

8. When George Sisler saw the figures “his reaction was one of bewilderment. ‘I still don’t believe it,’ he said. ‘But there it is’” (Rickey “Goodbye to Old Baseball Ideas,” 83).

9. 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.


11. 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.
From the Dugout to the Classroom

Why Good Baseball Coaches Have Much to Offer Good Professors

DARIN H. VAN TASSELL

Professors are many things. We are scholars who publish. We engage in service on our campuses and in our communities. But we are teachers first and foremost. We prepare young people to be successful in the classroom as well as in society. Despite what many see as an inherent tension on campus between faculty members and athletic coaches, those of us who teach students in the classroom have much to learn from those who coach student-athletes on the field. Yes, many athletic coaches receive more fame and fortune than most professors. And baseball coaches are certainly better known for their great pregame speeches, intimate knowledge of double-switch substitutions, and keen offensive strategies than professors are for their best lectures and finest articles. But the best coaches are also primarily teachers, and professors have much to learn from them. In many ways, baseball coaches—and perhaps most coaches—are much the same as the professor who teaches English, math, or history, for the best coaches use the teaching principles and skills used by the best classroom teachers.

THE TEACHING PHILOSOPHY OF GOOD BASEBALL COACHES

Let me explain. Armed with a solid knowledge of our respective disciplines, professors have information to share with our students. If we are interested in improving how we teach and educate our students, then we would be wise to borrow from the coaches’ playbook on how better to impart this disciplinary information to our students. For example, baseball coaches seek to understand how the body operates and how the mechanics of the sport should be executed. Thus, a coach can teach players what to do, how to do it, and most importantly why it should be done in that manner. All coaches teach what to do: “on this play, you go over there and field the ball.” Some