An Economic Evaluation of the *Moneyball* Hypothesis

Jahn K. Hakes and Raymond D. Sauer*
Clemson University

November 3, 2004

Note: This is version 1 of the paper. Since it has been linked to in various places, we preserve it here. We also note that a subsequent version which is lighter on productivity measurement issues, but more direct on the question of "how the A's did it" can be obtained here: http://hubcap.clemson.edu/~sauerr/moneyball-v2.pdf.

*Both Hakes and Sauer: The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, SC 29634-1309. Fax: 864-656-4192. Hakes: Phone: 864-656-5966; e-mail: jhakes@clemson.edu. Sauer: Phone: 864-656-3969; e-mail: sauerr@clemson.edu. We thank Mike Maloney and Bob Tollison for helpful comments.

An Economic Evaluation of the *Moneyball* Hypothesis

November 2004

Abstract

Michael Lewis's book, *Moneyball*, is the story of an innovative manager who exploits an inefficiency in baseball's labor market over a prolonged period of time. We evaluate this claim by applying standard econometric procedures to data on player productivity and compensation from 1999 to 2004. These methods support Lewis's argument that the valuation of different skills was inefficient in the early part of this period, and that this was profitably exploited by managers with the ability to generate and interpret statistical knowledge. This knowledge became increasingly dispersed across baseball teams during this period. Consistent with Lewis's story and economic reasoning, the spread of this knowledge is associated with the market correcting the original mis-pricing.

JEL codes: J31, O33, L83

Keywords: innovation, market correction, salaries and compensation, baseball

An Economic Evaluation of the *Moneyball* Hypothesis

Imagine a good produced primarily with labor. The particular skills involved are unique to the production of this good. The skills are also multi-dimensional, and individual workers have varying quantities of skill in each dimension. The associated labor market is well known. It receives daily attention from the print and broadcast media, along with periodic in-depth analysis from academic economists. Indeed, a case could be made that more is known about pay and performance in this market than any other labor market in the American economy.

Into this milieu strides a financial reporter who makes the following claim: the valuation of skill in this market is grossly inefficient. The inefficiency is so large that a former worker with mediocre talent and two quantitative analysts could exploit it to great effect. The firm that hired them was able to out-manage and out-produce the vast majority of the competition while operating on a shoestring budget. This is the tale told by Michael Lewis' (2003) instant classic, *Moneyball*, about the use of innovative thinking and statistical analysis in the management of the Oakland Athletics baseball club.

As the characterization above makes clear, the thesis of *Moneyball* is economic at its core, and indeed is potentially refutable. Yet "the biography of an idea," as Lewis (2004) called his work, does not verify itself, however convincing the argument. We examine the argument here. We cast it in refutable form, and test it with elementary econometric tools. We find that Lewis' claims bear close scrutiny. In particular, we confirm that the baseball labor market exhibited significant inefficiency in recent years.

This inefficiency was sufficiently large that knowledge of its existence, and the ability to exploit it, enabled the Athletics to obtain a substantial advantage over their competition. We also find that Lewis' timing was impeccable. The ideas in *Moneyball*, belying somewhat ignorant protestations from entrenched interests in the baseball world (Lewis, 2004), spread with sufficient speed that prices in baseball's labor market no longer exhibit the "*Moneyball* anomaly."

Sports often generate ideal conditions in which the choices of market participants can be observed and studied. The value of this is becoming more widely appreciated. Notable studies which illustrate the range of economic issues where data from sports can lend insight are Robert E. McCormick and Robert Tollison (1986), William O. Brown and Raymond D. Sauer (1993a, 1993b), and Pierre-Andre' Chiappori, Stephen D. Levitt, and Timothy Groseclose (2002). These papers analyze how the likelihood of punishment affects crime (fouls on the basketball court), the effects of psychology and information on market prices (point spreads for NBA games), and strategic optimization in a repeated game (the puzzling unwillingness of penalty-kick takers to shoot the ball down the middle, when goalkeepers almost always dive one way or the other).

The present paper's contribution may be thought of as a depiction of particularly clear case of mis-pricing, accompanied by profitable innovation and subsequent adjustment in the labor market. We document this by evaluating the measures of offensive productivity discussed in the book, and measuring their impact on game outcomes. We then assess the valuation of skill in baseball's labor market. Consistent with the claims made in *Moneyball*, important batting skills were undervalued in the marketplace during the initial periods that we study. However, a diffusion of the knowledge discussed in the book subsequently took place. We find that the anomaly disappears when members of the Athletics' organization were hired to run competing franchises.

The paper proceeds as follows. Section I describes measurements of batting skill in baseball. Section II relates these measures to the primary objective of winning games. It is here that we introduce the idea that certain elements of the game were not properly understood in the conventional wisdom of baseball. Section III examines the valuation of these skills in the labor market, and Section IV concludes.

I. Productivity Measurement in Major League Baseball

A. Standard Measures

Our focus here, as in Lewis (2003), is on the measurement of offensive productivity, or batting skill. The most common measure of skill is the *batting average,* i.e. the ratio of hits to total at-bats. The batting average is a crude index. By weighting singles and home runs the same, it ignores the added productivity from hits of more than a single base. Much better is the *slugging percentage* (total bases divided by at-bats) in which doubles count twice as much as singles, and home runs twice as much as doubles. Gerald Scully's (1974) demonstration that baseball players earned a small fraction of their marginal revenue product under the reserve system utilized slugging percentage as a measure of productivity. Nevertheless, both the batting average and slugging percentage ignore potentially relevant dimensions of batter productivity. Sacrifices and walks, for example, are often productive and occasionally crucial, but are ignored in both measures. Indeed, since a fundamental element of batting skill is the ability to avoid making an out, the failure to account for walks is a serious omission.[1] These flaws in the batting average were understood as early as the 1950s, when Branch Rickey[2] argued for the importance of "on base percentage," i.e. the fraction of at bats in which the player reached base successfully (Lewis, 2003; 71).[3]

The statistic du jour among statistically-minded followers of the game is "OPS," which is the sum of on base percentage (O) and slugging percentage (S). It has long been well known among this group, dubbed sabermetricians, that linear combinations of these two percentages are very highly correlated with runs scored, the primary objective of an offense.[4] The essence of the *Moneyball hypothesis* is that, although it was well known by sabermetricians that on base percentage was an important component of offensive productivity, the ability to get on base was seriously undervalued in the baseball labor market.

B. A Probability-Based Measure of Productivity

To assess the value of these measures of performance, we compare them with a more complicated, but conceptually superior alternative. In theory, an ideal measure of productivity would be tied directly to the primary objective – winning the game. This approach was pioneered in the academic literature by George R. Lindsey. Lindsey

---

[1] An inning in baseball is composed of three outs; once those are used up, any situational advantage derived from the number of men on base prior to that point is lost. Outs are the ultimate scarce resource in baseball, hence the ability of a batter to avoid them – which necessarily entails advancement towards home plate – is of fundamental importance.

[2] Rickey was general manager of the Brooklyn Dodgers in the 1940s, and is best known for breaking the color barrier in baseball by signing Jackie Robinson.

[3] Andrew Zimbalist's (1992) critique of the Scully model presciently noted that on base percentage makes an important contribution to statistical models of team winning percentage.

[4] The construction of OPS is similar to the "runs created" formula devised by Bill James (2001; 330), hence the high correlation between the two. James coined the moniker sabermetrics from the Society for American Baseball Research (SABR), of which James was a founding member.

(1961) analyzed the distribution of scoring throughout a game. Although scoring was non-stationary across innings, Lindsey found that a statistical model based on sampling from an independent and identical distribution of runs tracked the difference in the score of the game quite well. Lindsey concluded that his procedure provided a good estimate of "the probability that the game will eventually be won by the team that is ahead (p. 718)." In a subsequent paper, Lindsey (1963) used the same approach to conduct an illuminating analysis of managerial strategy.[5]

Lindsey's method provides the basis for Jay Bennett and J.A. Fleuck's (1984) concept of player game percentage (PGP). Bennett and Flueck measure player performance based on the impact of individual plays on the probability of winning, summed over all relevant plays. Events in PGPs are weighted in exact proportion to their impact on the probability of winning the game, and hence can be interpreted as fundamental measures of productivity. [6]

We use the Lindsey-Bennett-Fleuck method to estimate weights for computing a probability-based measure of batting productivity, or PGP. These weights are constructed by calculating the average impact of each event on the probability that a team wins the baseball game.[7] Fortunately, the most relevant events number in the thousands over an entire season, so the sample size for measuring the representative probability

---

[5] Lindsey's approach has been recently employed in empirical analyses of strategic choice in baseball. Turocy (2004) examines the strategic value of stolen base attempts, and Hakes and Sauer (2004) study sacrifice bunts, both adapting Lindsey's method to newly available play-by-play data.

[6] Since PGPs make use of detailed information about the state of the game before and after each play, they take into account potentially important information that is ignored in standard productivity estimates. Bennett's (1993) analysis of Shoeless Joe Jackson's performance in the infamous 1919 World Series – in which the timing of critical plays is of central importance - is perhaps the most notable application of the PGP measure of player productivity.

[7] Although state dependence is crucial for analyzing some questions, it is essential to net it out for an aggregate productivity measure. For example, the difference in the probability impact of a homer in the 9th inning of a tight game and a blowout can be 0.5 or higher, which can significantly affect productivity estimates for players with a modest number of at bats. We study this issue in Hakes and Sauer (2003).

change of an event is quite large. The appendix contains a detailed description of the method.

The estimated probability impacts for selected events are presented in Table 1. Both absolute and relative magnitudes appear quite sensible. As would be expected, the probability impact of a single (0.0418) outweighs that of a walk (0.0281) due to the possibilities for runner advancement, and extra base hits are worth more than singles. Similar events yield similar impacts, such as walks (0.0281), and batters hit by a pitch (0.0284). The effect of a batter grounding out into double play (GIDP) is more than three times the magnitude of a routine out, reflecting not only the loss of two outs, but also an advantageous position on the bases.

II. Statistical Evaluation of OPS and PGP Measures of Productivity

We now evaluate OPS and its component parts as measures of productivity, in comparison to the theoretically ideal PGP. For each team, we compute an aggregate OPS statistic based on all plate appearances during the season. Similarly, we calculate and aggregate a team-specific PGP-based measure, using as weights the mean impact of each event on win probability from Table 1. We label this statistic PGP*. The standard of comparison is the ability to explain differences in win percentage across teams between PGP* and OPS and its components. To incorporate the opposition's ability to score, we also construct measures of OPS and PGP achieved by a team's opponents throughout the season -- OPS against and PGP* against.

Table 2 reports coefficient estimates from regressing team winning percentage on our productivity measures for 1999 and 2000, the seasons for which we have the play-by-play data needed to calculate PGP* statistics. In both the OPS and PGP* models, the estimated contribution of productivity by the own team and its opponents are of similar magnitude, but opposite sign. While the PGP-based model has higher explanatory power (R^2 of .8926 vs. .8682), the performance of the OPS-model is surprisingly strong.[8] Breaking apart OPS into its on base and slugging components inches the explanatory power a bit closer to the PGP* model.

Why does OPS fare reasonably well when examined alongside PGP? Although it ignores relevant dimensions of productivity, the implicit weights in OPS are generally close to those obtained using optimal PGP weights. Table 1 shows that the average impact of a single on the probability of winning a baseball game is approximately 2/3 that of a double, and 1/3 that of a home run. These ratios are similar to the relative contributions of these events to the OPS measure. Hence the success of the OPS statistic is derived from that fact that the manner in which it weights the most common events is similar to the impact of these events on the probability of winning a baseball game.

The models in Table 2 are limited to the two seasons for which we can calculate PGP statistics. Section III's analysis of the labor market uses more readily available performance data over five seasons from 1999-2003. Table 3 makes use of this more extensive sample in estimating the impact of these productivity measures on winning percentage. The results closely match those in Table 2, with the OPS-based measures explaining 87% of the variance in winning percentage across teams.

---

[8] One can test whether one statistic adds information in the presence of the other. We fail to reject the hypothesis that OPS adds no information to PGP (p-value of .699). On the other hand, PGP* does add information to the regression with only OPS. By this standard, the linear weights-based PGP statistics are superior measures of productivity. But the level of effort required for PGP to modestly out-perform OPS suggests that the simple measure has its virtues.

The final column of Table 3 breaks down OPS into its components, and imposes the restriction that, loosely speaking, the contribution of Yankee bats to a Yankee victory is equivalent to the contribution of Yankee bats to their opponent's defeat. This model is used to assess a claim made in *Moneyball* (p. 128) that, contrary to conventional wisdom, on base percentage is more important than slugging percentage on a point-for-point basis. The coefficients in this regression are consistent with this claim: those for on base percentage are more than twice as large as the coefficients for slugging.

III. Valuation of Batting Skills in Baseball's Labor Market

In Scully's (1974) seminal model of pay and performance in baseball, a player's marginal revenue product is solely derived from his contribution to team winning percentage. The final column of Table 3 demonstrates that a one point change in team on base percentage makes a significantly larger contribution to team winning percentage than a one point change in team slugging percentage. As an individual player's attempts constitute similar shares of all team offensive percentage statistics, an efficient labor market would (ceteris paribus) reward on base percentage and slugging percentage in the same proportions that those statistics contribute to winning.

We estimate earnings equations for position players for the 2000-2004 seasons. The dependent variable is the logarithm of annual salary. All productivity variables are calculated based on performance in the prior year, and all players with more than 130 at bats in the previous season are included in the regressions.[9] The specification follows

[9] Since salary is determined prior to performance and is based on expected productivity, we use the prior year's performance as an indicator of expectations. A minimum of 130 at bats is required for a player to qualify for honors as rookie of the year. This provides an objective cutoff so that we employ productivity measures exclusively for players with a large sample of at bats.

that of Lawrence Kahn (1993), and includes indicator variables for labor market status. The base category is for younger players who have limited negotiating power under the collective bargaining agreement. Indicator variables reflect competitive bidding conditions for players eligible for arbitration and free agency, respectively. Other relevant control variables include playing time, as measured by plate appearances, and indicator variables for the more demanding defensive positions of catcher and infield. Following Kahn (1993), we define an infielder as a player with a primary defensive position at either second base, third base, or shortstop.[10]

The first column of results in Table 4 reports coefficient estimates from the log salary regression when all five years of data are pooled. All significant coefficients have the expected signs. There are large positive returns to contracting freedom. We estimate an incremental return to arbitration-eligible players relative to those subject to monopsony bargaining, and a still greater return for free agents. We also obtain positive and statistically significant returns to expected playing time. The returns to on base percentage and slugging are both positive, as expected. However, the coefficient for slugging is significantly greater than the coefficient for on base percentage, which is the reverse of their importance to team success. This is consistent with *Moneyball's* claim that on base percentage is undervalued in the labor market.

Columns 2 through 6 of Table 4 display parameter estimates for the same equation for each individual season. These results indicate that pooling is inappropriate, as labor market returns to player attributes appear to differ across seasons. This is clearly

---

[10] Productivity and positional data were obtained from the Lahman baseball database at the Baseball Archive, http://baseball1.com. Data on salaries and labor market status were obtained from Doug Pappas' Business of Baseball data archive, http://roadsidephotos.sabr.org/baseball/data.htm.

the case for the estimated returns to on base percentage and slugging percentage, as Figure 1 illustrates.

In the first four years of data, the slugging coefficients are all statistically significant and of similar magnitude, ranging between 2.05 and 3.10. In contrast, the on base percentage coefficients are smaller than their slugging counterparts (between –0.13 and 1.36) in each of these years and are not significantly related to log salary. The lack of a market premium for hitters with superior skill at reaching base validates the Athletic's systematic approach to identifying such players, and thereby winning games at a discount relative to their competition.[11]

The relative valuation of on base and slugging percentage is abruptly reversed for the year 2004. The returns to slugging are similar in 2004 to prior years, but this is the first year for which the ability to reach base is statistically significant. The labor market in 2004 appears to have substantially corrected the apparent inefficiency in prior years, as the coefficient of on base percentage jumps to 3.68, and the ratio of the monetary returns to reaching base and slugging is very close to the ratio of the statistics' contributions to team win percentage.

IV. Concluding Remarks

Our analysis supports the hypothesis that baseball's labor market was inefficient at the turn of the twenty-first century. Exploitation of this inefficiency by the Oakland Athletics' organization is traced by Lewis (2003; 58-63) to an explicit decision, inspired

[11] As discussed in Lewis (2003, xii), Doug Pappas (chairman of SABR's Business of Baseball Committee), calculated the incremental cost of of winning a game during this period. Only two teams spent less than $1m to win a game. The Athletics' cost of about half a million dollars was the lowest, and about 1/6 the cost of the least efficient teams. Pappas (2002) discusses the calculation and provides costs estimates for all teams during the 2001 season.

by the work of Bill James, to fuse statistical analysis of the game into a management strategy. To execute this strategy, the A's reached outside baseball circles and hired talented young analysts with Ivy League pedigrees and an interest in the game.

The particular margin of inefficiency emphasized in Moneyball, undervaluation of a batter's ability to get on base, appears to have been substantially if not completely eroded by the time the book was published. Despite protests from baseball traditionalists (Lewis, 2004) that the book was somehow misguided, several major league teams had by this time decided to imitate the strategy.

During and after the 2003 season, two young analysts from the Athletics' front office were hired as General Managers by the Toronto Blue Jays and the Los Angeles Dodgers (Joe Saraceno, 2004).[12] Although the Boston Red Sox failed in their attempt to hire both the Athletics' General Manager (Billy Beane) and Assistant GM, they followed Beane's advice by hiring Theo Epstein, making him the youngest GM in baseball history (Shaughnessy, 2003). In addition, the Sox hired the father of sabermetrics, Bill James himself, in an advisory capacity. This diffusion of statistical knowledge across a handful of decision-making units in baseball was apparently sufficient to correct the mis-pricing of skill.

Finally, we close on a humble note. Our work is essentially an assessment of Lewis' argument, and as such is merely an after-the-fact replication of work done by the innovators at the heart of his book. But we do find that they (and Lewis) were right, and further, that the process that they set in motion had in large part been completed by the time the book was published. That the addition of only a few individuals was able to correct a long-standing anomaly illustrates both the inefficiency which can result when

[12] For most franchises, the General Manager is the executive with authority over personnel decisions.

markets are isolated from competition, and the swiftness of market corrections once entry occurs.

Table 1: Average Effect of Selected Events Upon Probability of Winning

Event	Frequency	Mean Change in P(win)	Std. Error
Walk	17,028	0.0281	0.0002
Hit by pitch	1,572	0.0284	0.0006
Single	29,686	0.0418	0.0003
Double	8,902	0.0646	0.0007
Triple	952	0.0948	0.0026
Home run	5,693	0.1217	0.0013
Strikeout	31,254	-0.0276	0.0001
Ground out	35,191	-0.0220	0.0001
Fly out	25,279	-0.0248	0.0001
Ground into DP	3,833	-0.0753	0.0010

Notes: The table reports the average change in the probability of winning associated with each event across all game states, along with the frequency of each event and the standard deviation of the mean probability. Based on scoring probabilities from each game state, using all plays from the 1999 season. Data Source: Stats Inc.

Table 2 – Predicting Winning Percentage: PGP vs. OPS

	Model		
	1	2	3
Constant	0.500	.483	.400
	(0.003)	(0.102)	(0.124)
PGP*	0.0071		
	(0.0005)		
PGP* Against	0.0069		
	(0.0005)		
OPS		1.325	
		(0.097)	
OPS Against		(-1.304)	
		(0.092)	
On Base			2.0060
			(0.3849)
On Base Against			-1.5539
			(0.3768)
Slugging			0.9898
			(0.2064)
Slugging Against			-1.1179
			(0.2227)
Number of Obs	60	60	60
R^2	.8926	.8682	.8768

Notes: The dependent variable is team win percentage, for years 1999-2000. The data are season level aggregates. PGP* statistics constructed from situation-specific play-by-play data for every game of the 1999 and 2000 seasons. We have only two years of play-by-play data for comparison, hence the reduced number of observations. Source: Stats Inc. and ESPN.com. Standard errors are in parentheses.

Table 3 – Productivity Estimates:
The Impact of On Base and Slugging Percentage on Winning

	Model		
	1	2	3
Constant	0.558	0.508	.500
	(0.073)	(0.073)	(0.002)
OPS	1.261		
	(0.061)		
OPS Against	-1.337		
	(0.055)		
On Base		2.121	2.032
		(0.255)	(0.180)
On Base Against		-1.912	-2.032^R
		(0.256)	
Slugging		0.812	0.900
		(0.141)	(0.105)
Slugging Against		-0.995	-0.900^R
		(0.144)	
Number of Observations	150	150	150
R^2	.871	.885	.882

Hypothesis Tests
Model 2, H^0:
On Base = -On Base Against
Slugging = -Slugging Against
F(2,145) = 0.49, p-value = 0.613
Model 3, H^0:
On Base = Slugging
F(1,147) = 17.21, p-value = 0.0001

Notes: Data are aggregate statistics for all 30 teams from 1999-2003. Coefficient estimates obtained using OLS. Standard errors are in parentheses; ***, **, and * mean that the coefficient is statistically different from zero at the 1-, 5-, and 10-percent level, respectively. The superscript R indicates the coefficient was restricted to equal its counterpart in the regression.

Table 4 – The Baseball Labor Market's Valuation of On Base and Slugging Percentage

	All Years	2000	2001	2002	2003	2004
On Base	1.360	1.334	-0.132	0.965	1.351	3.681
	(0.625)	(1.237)	(1.230)	(1.489)	(1.596)	(1.598)
Slugging	2.392	2.754	3.102	2.080	2.047	2.175
	(0.311)	(0.628)	(0.613)	(0.686)	(0.850)	(0.788)
Plate Appearances	0.003	0.003	0.003	0.003	0.003	0.003
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Arbitration Eligible	1.255	1.293	1.106	1.323	1.249	1.323
	(0.047)	(0.102)	(0.100)	(0.100)	(0.111)	(0.115)
Free Agency	1.683	1.764	1.684	1.729	1.663	1.575
	(0.044)	(0.096)	(0.092)	(0.097)	(0.107)	(0.105)
Catcher Dummy	0.152	0.137	0.065	0.208	0.343	0.059
	(0.056)	(0.124)	(0.116)	(0.122)	(0.134)	(0.133)
Infielder Dummy	-0.029	0.060	0.069	-0.087	-0.054	-0.100
	(0.040)	(0.087)	(0.083)	(0.086)	(0.095)	(0.098)
Intercept	10.083	10.078	10.347	10.490	10.289	9.782
	(0.170)	(0.360)	(0.321)	(0.358)	(0.387)	(0.414)
Number of observations	1736	353	357	344	342	340
R^2	0.675	0.676	0.728	0.695	0.655	0.635

Notes: The dependent variable is ln Salary for year t, and performance variables are from year t-1. 0/1 dummies for each year are included in the pooled regression. Standard errors in parentheses. The sample includes all players with at least 130 plate appearances during the relevant season.

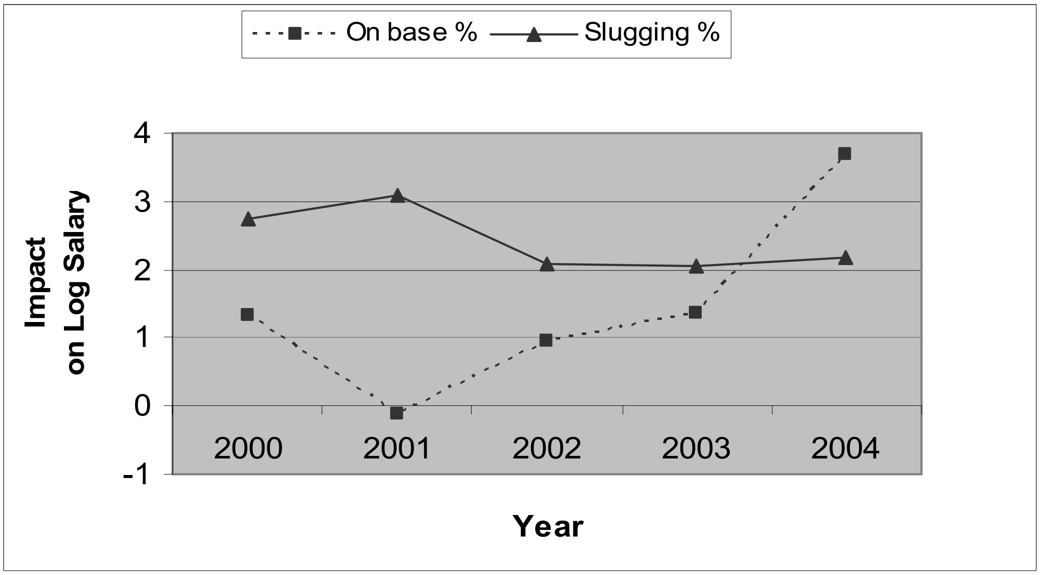

Figure 1. Labor Market Returns to On Base and Slugging Percentage Over Time

APPENDIX

Measuring the Impact of Plays on the Probability of Winning

To measure the impact of a play, we must construct estimates of the probability of winning before and after the play. This requires knowledge of the conditional probability of scoring in each of the various possible game situations, or states. The state of the game in this context means the number of outs and the number and location of men on base. Table A1 describes the conditional probability distribution of scoring in each possible state. The distributions were constructed from play-by-play data obtained from Stats Inc., consisting of over 190,000 observations from the 1999 season. Each of the 24 possible states contains at least 524 observations. Moving down the first column, we see that the probability of not scoring any runs, conditional on there being no men on base, increases from 0.695 to 0.918 as the number of outs increases from 0 to 2. In the far right column we see that the expected number of runs scored in an inning diminishes from 0.577 to 0.124 as we make this same progression from 0 to 2 outs with no men on base.

Using backward induction and the probabilities in Table A1, we can calculate the probability of winning in each inning, conditional on the run difference, basecode and number of outs. Let P_H (h,I,b,o,d) represent the probability that the home team, H, wins a game situated in the home half (h) of inning I, with runners indicated by basecode b, o outs, and facing a run difference of d runs. If the home team is trailing at the start of the bottom of the ninth inning, the probability that it will win is

$$P_H(1,9,0,0,d) = P_S(R>d|0,0) + .5* P_S(R=d|0,0)$$

where $P_S(R|b,o)$ is the stationary probability function for scoring R runs during the inning conditional on basecode and outs, with a score difference (runs less opponents runs) of d runs at the start of the ninth. At the start of the ninth, the conditional probabilities are taken from the first row of Table A1. As outs are recorded and/or runners advance, the probability is updated to match the new state of the game. The relevant columns for the probability sums also change when scoring alters the run differential d.

Once we know the probability that the home team overcomes a deficit in the bottom of the ninth, we can take any run difference facing the visiting team in the top of the ninth, the probability it scores and hence changes the run difference, and the probability that the home team overcomes this new difference (if necessary), and thereby compute the probability that the visiting team is victorious given any situation it faces in the top of the ninth. For example,

$$P_V(0,9,0,0,d) = \sum_{R=0}^{9} P_S(R \mid 0,0)(1 - P_H(1,9,0,0,-(d+R)))$$

where $P_V(h,I,b,o,d)$ is the visitor's probability of winning the game.[13] In recursive fashion, the probabilities can be computed in this manner all the way to the start of the game.

These computed probabilities have two shortcomings, but we do not believe they compromise our approach. First, batting order and endgame effects could render scoring non-stationary, so that the probabilities from Table A1 would not be appropriate for use in all situations. As batting orders are ideally set up at the start of the game, scoring might be greater in the first inning than in other innings. Similarly, strategic choices by managers may alter scoring distributions in the late innings of close games. Second, the quality of the batting lineup and the pitchers will also affect the numbers in the table. But

[13] The summation theoretically should run from 0 to infinity, but we use a maximum of nine runs scored in a half inning. Innings of this magnitude are extremely rare, and result in large score differences where the probability of winning approaches 1.00. This simplification increases computational efficiency while having a negligible effect on the estimates.

in both cases, it is likely that these effects will impact all numbers by a similar magnitude. Since our measure of performance, PGP, is defined as the difference between the win probabilities before and after the play, the differencing of any bias of constant magnitude will result in an unbiased PGP estimate.[14]

The sum of the team's PGP values for all events that take place while it is on offense during the season – outs, walks, sacrifice flies, home runs, etc. – represents the team's offensive productivity. The third column of Table A2 presents the unadjusted PGP totals for each team for the 1999 season. The fourth column, labeled PGP*, weights each event by its mean change in probability as presented in Table 1, which removes the effects of the timeliness of the event. Teams are ranked in Table A2 by PGP*. The difference between offensive team PGP and PGP* may be due either to differing productivity in "clutch" situations, or it may be just luck.[15]

For purposes of comparison, the fifth column of Table A2 includes each team's OPS statistic, as OPS is the most common single measure of offensive productivity in use today.[16] The correlation coefficient between OPS and PGP* is very high at 0.9923. The rightmost column of Table A2 displays team PGP* values normalized to the same scale as OPS.[17] The difference between the two measures is quite small in all cases. PGP* is more highly correlated with team success, as Table 2 demonstrates. But the close correspondence between OPS and the theoretically superior measure of PGP* confirms that, OPS is a remarkably good index of productivity.

---

[14] A standard χ^2 test rejects the null hypothesis that the expected number of runs is stationary across innings ($\chi^2_9 = 166.4$; $p = 0.000$) due to these reasons. Non-stationarity in the probability of scoring and the effect of high(low) scoring teams upon scoring distributions remain issues that require further exploration.
[15] We explore the nature of clutch performance in Hakes and Sauer (2003).
[16] Team OPS is calculated in the same manner as an individual player's OPS, pooling the results of all plate appearances by all batters on the team.
[17] The normalization is based on coefficients from the OLS regression: OPS = 0.791 + 0.0055 PGP* ($R^2 = 0.985$).

Table A1: Probability of Runs Scored in an Inning, by Basecode and Outs

Base code	Outs	p(0)	p(1)	p(2)	p(3)	p(4)	p(5+)	E(runs)	Obs
0	0	0.695	0.159	0.078	0.038	0.017	0.014	0.577	45495
0	1	0.818	0.105	0.046	0.019	0.007	0.006	0.313	31968
0	2	0.918	0.054	0.019	0.006	0.002	0.001	0.124	25392
1	0	0.558	0.168	0.138	0.071	0.035	0.031	0.972	10804
1	1	0.709	0.117	0.097	0.044	0.018	0.016	0.600	12227
1	2	0.858	0.058	0.056	0.019	0.005	0.004	0.267	11946
2	0	0.368	0.352	0.139	0.074	0.040	0.027	1.170	3470
2	1	0.585	0.236	0.098	0.050	0.017	0.014	0.727	5867
2	2	0.780	0.147	0.047	0.017	0.005	0.003	0.329	7448
3	0	0.136	0.508	0.189	0.105	0.025	0.038	1.513	524
3	1	0.329	0.485	0.111	0.044	0.022	0.010	0.980	2010
3	2	0.731	0.186	0.054	0.018	0.006	0.004	0.398	3054
4	0	0.349	0.218	0.164	0.130	0.071	0.068	1.616	2886
4	1	0.569	0.160	0.104	0.094	0.039	0.034	0.998	5123
4	2	0.760	0.108	0.060	0.048	0.016	0.008	0.479	6435
5	0	0.116	0.435	0.165	0.136	0.082	0.067	1.889	1068
5	1	0.340	0.380	0.118	0.091	0.044	0.027	1.214	2277
5	2	0.729	0.146	0.054	0.051	0.014	0.006	0.496	2886
6	0	0.154	0.246	0.308	0.138	0.072	0.082	2.034	668
6	1	0.291	0.296	0.212	0.101	0.054	0.045	1.490	1695
6	2	0.727	0.048	0.146	0.047	0.017	0.016	0.628	1911
7	0	0.114	0.257	0.208	0.120	0.153	0.148	2.509	802
7	1	0.315	0.260	0.141	0.112	0.094	0.078	1.691	1949
7	2	0.672	0.091	0.107	0.055	0.049	0.025	0.807	2356

Notes: Data Source: Stats Inc. Play-by-Play data for 1999. Basecodes: 1 = runner on 1^{st}; 2 = runner on 2^{nd}; 3 = runner on 3^{rd}; 4 = runners on 1^{st} and 2^{nd}; 5 = runners on 1^{st} and 3^{rd}; 6 = runners on 2^{nd} and 3^{rd}; 7 = bases loaded. Calculations extend to four decimal places, allow for scoring of up to 9 runs in an inning, and track p(win) for run differentials of (+/-) 9 runs.

Table A2: Offensive PGP Totals and OPS for 1999

Rank	Team	PGP	PGP*	OPS	Normed PGP*
1	CLE	15.603	8.865	0.840	0.840
2	TEX	10.359	7.565	0.840	0.833
3	NYY	5.420	4.614	0.819	0.816
4	COL	7.047	4.379	0.819	0.815
5	ARI	5.162	2.831	0.805	0.807
6	TOR	4.038	2.735	0.810	0.806
7	OAK	5.377	2.268	0.801	0.804
8	BAL	-1.739	2.117	0.800	0.803
9	BOS	-0.253	1.384	0.798	0.799
10	SEA	1.600	1.255	0.798	0.798
11	NYM	2.165	1.077	0.797	0.797
12	CIN	1.938	0.923	0.792	0.796
13	SF_	-1.088	0.549	0.790	0.794
14	ATL	-0.536	-1.271	0.777	0.784
15	HOU	-4.411	-1.562	0.775	0.783
16	KC_	-3.536	-2.111	0.781	0.780
17	PHI	-5.745	-2.325	0.782	0.778
18	MIL	-1.756	-2.877	0.779	0.775
19	STL	-4.437	-3.849	0.764	0.770
20	CWS	-3.409	-4.030	0.766	0.769
21	LA_	-9.204	-4.960	0.760	0.764
22	DET	-8.696	-5.271	0.768	0.762
23	PIT	-9.882	-7.280	0.753	0.751
24	CHC	-3.259	-8.126	0.749	0.747
25	TB_	-9.719	-8.274	0.754	0.746
26	MON	-9.762	-8.335	0.751	0.746
27	SD_	-14.152	-11.457	0.725	0.728
28	FLA	-15.741	-13.601	0.719	0.717
29	MIN	-16.087	-13.624	0.712	0.717
30	ANA	-13.018	-14.050	0.716	0.714

Notes: Teams are ranked by PGP*. Normed PGP* is PGP* normalized to the scale of OPS, using OLS regression coefficients.

References

Bennett, Jay. "Did Shoeless Joe Jackson Throw the 1919 World Series?" *The American Statistician*, November 1993, 47: 241-250.

Bennett, Jay, and Fleuck, J. A. "Player Game Percentage," In *Proceedings of the Social Statistics Section, American Statistical Association*, 1984, 378-380.

Brown, William O. and Sauer, Raymond D. "Fundamentals or Noise? Evidence from the Basketball Betting Market," *Journal of Finance*, September 1993a. 48(4): 1193-1209.

Brown, William O. and Sauer, Raymond D. "Does the Basketball Market Believe in the Hot Hand: Comment," *American Economic Review*, December 1993b. 83(5): 1377-86.

Chiappori, Pierre-Andre', Levitt, Steven, and Groseclose, Timothy. "A Test of Mixed Strategy Equilibria: Penalty Kicks in Soccer," American Economic Review, September 2002. 92: 1138-1151.

Hakes, Jahn K., and Sauer, Raymond D. "Are Players Paid for Clutch Performance?" manuscript, July 2003, Clemson University.

Hakes, Jahn K., and Sauer, Raymond D. "Measurement and Evaluation of Managerial Efficiency in Major League Baseball," manuscript, June 2004, Clemson University.

James, Bill. *The New Bill James Historical Baseball Abstract*, Free Press, 2001: New York.

Kahn, Lawrence M. "Free Agency, Long-Term Contracts and Compensation in Major League Baseball: Estimates from Panel Data," *The Review of Economics and Statistics*, February 1993, 75: 157-164.

Lewis, Michael. *Moneyball: The Art of Winning and Unfair Game*, Norton, 2003: New York.

Lewis, Michael, "Out of Their Tree," *Sports Illustrated*, March 1, 2004.

Lindsey, George R. "The Progress of the Score During a Baseball Game," *Journal of the American Statistical Association*, September 1961. 56(295): 703-728.

Lindsey, George R. "An Investigation of Strategies in Baseball," *Operations Research*, July-August 1963. 11(4): 477-501.

McCormick, Robert E., and Tollison, Robert D. “Crime on the Court,” *Journal of Political Economy*, April 1984, 92(2): 223-35.

Pappas, Doug. "The Numbers (Part Four): Player Compensation," http://www.baseballprospectus.com/article.php?articleid=1320.

Saraceno, Joe. "Dodgers Turn to Ivy League," *USATODAY*, March 17, 2004.

Scully, Gerald P. "Pay and Performance in Major League Baseball," *American Economic Review*, December 1974, 64: 915-30.

Shaughnessy, Dan. "Beane Has Looked Sharp By Doing Things His Way," *Boston Globe*, September 28, 2003.

Turocy, Theodore L. "A Theory of Theft: The Strategic Role of the Stolen Base in Baseball," manuscript, January 2004, Texas A&M University.

Zimbalist, Andrew. "Salaries and Performance: Beyond the Scully Model," In *Diamonds are Forever*, P. Sommers ed., Brookings, 1992: Washington D.C.