A Probability Based Measure of Productivity in Major League Baseball With Application to the Questions of Clutch Performance & the Value of Pitching

Jahn K. Hakes

and

Raymond D. Sauer

April 22, 2004

John E. Walker Dept. of Economics, Clemson University

Preliminary And Incomplete: Comments Welcome

E-mail: jhakes@clemson.edu; sauerr@clemson.edu.
Mail: 222 Sirrine Hall, Box 1309, John E. Walker Dept. of Economics, Clemson University, Clemson, SC 29634-1309.

Foreword to the ASU Seminar

The main body of this paper presents and evaluates a probability-based measure of productivity in major league baseball, and applies this measure to the question of clutch performance. These measures make use of data on every play throughout the season. The concept is based on the effect of a player's action - say hitting a two run homer or grounding into a double play - in changing the state of the game, and thus increasing (or decreasing) his teams' probability of winning the game.

We apply this measure of productivity to the question of clutch performance. Similar to the "hot hand" issue, there is a debate over the existence of clutch hitting among major league players. Most statistical measures of performance show no persistence in clutch hitting; hence belief in clutch performance may represent another manifestation of mistaken belief in persistence in the presence of truly uncorrelated random sequences. It may be however, that standard measures fail to detect the existence of clutch hitting because they ignore productive plays (i.e. a sacrifice fly, or even "hitting behind the runner") which players may strategically pursue, given the state of the game. A probability-based measure stands a better chance of detecting persistence in clutch performance should these types of plays be common and important.

We find that our measures improve upon standard measures of productivity, in the sense that they are superior predictors of team success. However, in the limited sample used in this paper, our measures fail to detect the existence of clutch performers. However, we find some evidence that the labor market rewards the *appearance* of clutch performance. Hence we are faced with a mystery, one which we have not yet solved. To do so we are expanding our sample to more than a decade of play-by-play data. This project is motivated by the hope that a number of behavioral and strategic issues in economics (in addition to clutch effects per se) can be fruitfully addressed with a large sample of data encompassing strategic decisions made by participants in this setting.

The value of using data from sports to address economic questions is becoming more widely appreciated. Sports present repeated situations in which participants make strategic choices, generating useful data on these choices and outcomes. Notable examples of papers using data from sporting contests are McCormick and Tollison (1986), Chiappori, Levitt, and Groseclose (2002), and Romer (2003). These papers address a range of topics, from how the probability of punishment affects crime (basketball), to strategic choices in kicking situations (penalty kicks in soccer, punting in American football).

The plan for the seminar is first, to introduce our measure of productivity, and evaluate it relative to commonly used measures. Two applications will follow. First, we consider the issue of clutch performance. All of this material is fully covered in this paper. I will also discuss a method for estimating the value of starting pitching in baseball. This method is also probability-based, and incorporates information available from the baseball betting market. This latter application is not integrated into the main body of the paper. A sketch of the empirical model based on the betting market, along with some parameter estimates, is included in the appendix to this paper. This section, though sketchy, is an important component of where this paper is headed.

I. Introduction

In this paper we consider the possibility that the labor market rewards something that does not exist -- "clutch" performance. Although important achievements on the field are perpetually discussed among sports fans, the nature of clutch performance has received little attention in economic and statistical research. To date, we know only of work by "sabermetricians" (i.e. baseball aficionados armed with statistical skills) who generally conclude that the phenomenon of clutch performance is a myth (Neyer, 1999). Nevertheless, legends are made by players who succeed when games are on the line. In what follows, we provide a thorough investigation into the nature and valuation of these feats.

We construct a new measure of offensive productivity because assessments of clutch performance based on traditional measures are limiting. For example, Brooks (1989) cast doubt on an earlier claim of detecting clutch effects using batting average. Brooks showed that batting average "in the clutch" -- in the sense of deviation from normal situations -- was uncorrelated across seasons. But it is well known among sabremetricians that batting average, while popular, is a flawed measure of productivity -- it ignores many acts of contributing to team success. Earlier, Cramer (1977) reached a similar result using formulas which attach various weights to outcomes realized in each plate appearance. Our approach is similar to Cramer's. The principal difference lies in our exclusive reliance on probability theory to determine the weights attached to each outcome.

The basis for our approach is Bennett and Fleuck's (1984) concept of player game percentage (PGP). Bennett and Flueck measure player performance based on the impact of particular plays on the probability of winning a game.[1] Since PGPs make use of detailed information about the state of the game when each play is made, they take into account potentially important information that is ignored in standard productivity estimates.

PGPs are particularly well suited for analysis of clutch effects, as indicated by Bennett's (1993) analysis of Shoeless Joe Jackson's performance in the 1919 World Series.[2] The conceptual basis for PGP is the impact of a particular event on the probability of winning a game. Hence, a late inning single in a tight game receives more credit than a double in a game that is far out of reach. If one wants to seriously examine the possibility that clutch effects exist, this manner of allocating credit should be considered. Furthermore, whereas standard productivity measures like the "batting

---

[1] In Hakes and Sauer (2002) we focus on measuring the contribution of fielding skills with these methods.

[2] Bennett used PGPs to assess the likelihood that Jackson purposefully played to lose the 1919 World Series. Bennett showed that Jackson was quite productive when the game was on the line and that unlike his other "Black Sox" brethren, Jackson's performance in the Series materially increased his team's opportunity to win. If Shoeless Joe was cheating, he apparently did a bad job of it.

average" are arbitrary in nature, PGPs are closely tied to the main objective: winning the game. 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.

The existence of clutch effects, like that of the hot hand, is controversial. Legendary feats in crucial, championship-winning circumstances may be the root cause of belief in clutch effects. Such exhibitions of grace under pressure may create beliefs in clutch performance that are disproportionate to reality.[3] No one doubts that major league baseball players, as a rule, are capable performers in pressure situations. But for clutch performance to have statistical meaning, it must be shown to be non-random, i.e. in some way predictable. Hence, the question is whether pressure situations systematically change relative productivity among players -- that is, whether some players consistently perform better in the clutch than other players of similar ability in "normal" circumstances.

As Neyer (1999) discusses, the limited evidence in favor of clutch effects is quite weak. However, in a fashion similar to the hot hand, the power of statistical procedures employed thus far may be insufficient to distinguish between the views of "it's clutch" and "it's luck" partisans.[4] We seek to determine whether construction of PGP measures of performance can shed further light on this issue.

We show below that the measure of productivity we develop is superior to standard measures used to evaluate player performance. These standard measures are employed in the economic literature to assess the relation between player compensation and productivity. Generally speaking, studies show that in the free agent era, eligible players earn their marginal revenue products. But these studies use performance measures which can be improved upon. The economic question we address is whether players receive compensation in accord with perceptions of productivity -- clutch included -- which may have no relation to predicting the outcome of games. Recently, the publication of Lewis' (2003) *Moneyball* has gained considerable attention. Lewis makes the claim that baseball team management, with a few notable exceptions, remains stuck in the dark ages when it comes to performance assessment. Significant inefficiencies arise because pay is not appropriately linked to productivity. Our methods enable a systematic assessment of this claim.

The paper proceeds as follows. Section II discusses the method used to create PGP measures of offensive productivity, and compares our PGP-based measure of productivity with alternative measures that are currently employed. In Section III we

[3] Kirk Gibson's improbable 9th inning home run in the 1988 World Series, and Michael Jordan's buzzer-beating jumper to win the 1998 NBA Championship come to mind.

[4] Brown and Sauer (1993) explore the hot hand issue in the context of team streakiness in the NBA and beliefs as manifested in the betting market. They are able to provide definitive evidence of the market's belief in the hot hand. The noisiness of game outcomes is sufficient however to accommodate both believers and skeptics. The null hypothesis that there is no such thing as streaky performance among NBA teams is not rejected. However, the hypothesis that streak effects exist -- of the magnitude perceived by the betting market -- is also consistent with the data.

analyze the existence of clutch performance as measured by our PGP-based method. Section IV studies the valuation of random acts of performance. Section V concludes.

II. Construction of Player Game Percentage Based Estimates of Offensive Productivity

A. A Critique of Commonly Employed Productivity Measures

Commonly employed measures of hitting prowess such as a player's batting average or slugging percentage are essentially weighted averages based on arbitrary weights. The batting average, for example, is a zero/one scheme in which all hits -- from singles to home runs -- receive a value of one, and all failed at-bats receive a value of zero. Sacrifices and walks, which are often productive and occasionally crucial, are simply ignored in the batting average. The "on base percentage" statistic improves things by taking account of walks, but still ignores potentially productive sacrifices.

In the slugging percentage (total bases divided by total at-bats), doubles receive twice the weight of singles, and homers twice the weight of doubles, but again, these weights are somewhat arbitrary. Recently, the concept of "OPS" -- on base plus slugging percentage -- has become a widely used index of batting productivity. The OPS statistic takes into account a players' propensity to get on base -- in particular, to avoid making an out -- in addition to his ability to hit for extra bases. However, although credit for walks is given, the weights in this statistic are just as arbitrary as those in the slugging percentage. Our strategy is to use data from the evolution of play in baseball games to inform us of the optimal weights to attach to each outcome when constructing a single index of offensive productivity.

We use Bennett's method to estimate weights for use in computing a modified batting average or slugging percentage statistic. These weights are constructed by calculating the average impact of each event on the probability that a team wins the baseball game. It turns out that -- while the OPS statistic ignores relevant dimensions of productivity -- its implicit weights are generally close to those obtained by calculating optimal PGP weights. We find, for example, that the average impact of a single on the probability of winning a baseball game is approximately 2/3 that of a double, which is similar to the relative weights attached to the two events in OPS.

B. 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 various situations, or states of the game. The state of the game in this context means the number of outs and the number and location of men on base. The conditional probability of scoring in these various states is shown in Table 1. Table 1 is based on analysis of pay-by-play data from Stats Inc. for the 1999 season. 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 .577 to .124 as we make this same progression from 0 to 2 outs with no men on base.

Our calculations would be greatly simplified if the distribution of scoring is stationary across innings. Since batting orders are ideally set up at the start of the game, scoring might be greater in the first inning than in other innings. However, a standard Chi-squared test fails to reject the null hypothesis that the expected number of runs is stationary across innings (test stat, p-value).

Using the probabilities in Table 1, and the assumption of stationarity, we use backward induction to 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 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 wins 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 1. As outs are recorded and/or runners advance, the probabilities move to the row which matches the basecode and out situation of the game. The relevant columns for the probability sums also change when d changes as the home team scores.

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. In recursive fashion, the probabilities can be computed in this manner all the way to the start of the game.

Table 2 contains a selected group of probability estimates generated in this way to illustrate the relative impact of changes in the state of the game at various stages. Section A of Table 2 indicates the increasing value of plays as the game wears on. The value of a proficient leadoff hitter (such as Rickey Henderson, the record holder for career leadoff home runs) is apparent in the second row of the table. Nevertheless, a single or one-base error[5] in the ninth inning of a tight game is worth about twice that of a single in the first inning. More pronounced is the ninth inning home run, which is worth about three times as much as its first inning counterpart. Timing matters.

Section B of Table 2 illustrates how the value of a small (one run) and not-so-small (three run) lead increases as the game moves towards its conclusion. Section C measures the probability of winning at the start of the bottom half of innings 1 and 9 in a tie game. The difference for the home team of 5 and 15 percentage points (over the $P_V(0,I,0,0,0)$ of .500 at the top of the inning) indicates the increasing value to the home team of recording a scoreless top half of the inning.

There are two shortcomings of these probabilities, but we do not believe they compromise our approach. First, endgame and batting order effects are likely to make scoring probabilities non-stationary. Second, we have ignored the identity of the home team and the pitchers, which will also affect the numbers in the table. But these effects will impact all numbers by a similar magnitude. Since our measure of performance is based on changes in probabilities, differencing any bias of constant magnitude will result in an unbiased probability estimate. Nevertheless, non-stationarity in the probability of scoring remains an issue that requires further exploration.

C. PGP Based Measures of Productivity

Using the scoring probabilities from Table 1, we calculate the change in probability of winning for every event, i.e. single, walk, sacrifice fly, etc. This change in probability is, of course state dependent. Although state dependence is crucial for analyzing clutch performance, it is essential to we net it out for an aggregate productivity measure. We do this by calculating the average change in probability that occurs when a particular event takes place. Fortunately, the number of such events over an entire season is often in the thousands, so the sample sizes for measuring the representative probability change of event is quite large.

These estimates are presented in Table 3. As would be expected, the probability impact of a single (.0418) outweighs that of a walk (.0281) due to the possibilities for runner advancement, and extra base hits are worth more than singles. Similar events such as walks and batters hit-by-pitch yield similar impacts (.0281 and .0284). The negative effect of a ground into double play (GIDP) on the offensive team's probability of winning is more than three times the magnitude of a routine ground out, and errors are only very slightly less costly to a team than allowing a hit for an identical number of bases.

We now construct team measures of PGP-based productivity for offense. For each event that takes place in the season -- an out, walk, sacrifice fly, home run, etc. -- the impact of that event is added to a team's total PGP. Table 4 presents PGP totals for each team for the 1999 season. PGP is the summation of the unadjusted changes in probability associated with each play that occurred through out the season. LWPGP weights each event by the associated average change in probability from Table 3. Teams

---

[5] More precisely, anything that changes the state from bases empty to having a man on first.

are ranked in Table 4 by LWPGP. Teams with PGP higher than LWPGP are either clutch teams, or just lucky.

For purposes of comparison, Table 4 includes each team's OPS statistic.[6] OPS is the standard single measure of productivity in use today. The correlation coefficient between OPS and LWPGP is very high at .9923. A PGP measure normalized to the same scale as OPS is included in the rightmost column of Table 4.[7] The difference between the two measures is quite small in all cases. The close correspondence between OPS and LWPGP confirms that OPS, given its simplicity, is a remarkably good index of productivity. It also suggests that the extra effort involved in calculating PGPs might be futile!

Nevertheless there are reasons to expect that PGPs are more closely related to team success than OPS. First, recall OPS simply adds together a batter's on base percentage and his slugging percentage. The implicit weights used in OPS are arbitrary, whereas the PGP weights are determined by the empirical contribution of each event to the probability of winning a ballgame. Second, some contributions -- i.e. a crucial sacrifice fly -- are simply ignored by OPS. Finally, OPS completely ignores timing, which is crucial for winning and also evaluation of clutch performance. To the extent that factors omitted by OPS are important or the relative weights erroneous, the PGP statistic may be a superior measure of productivity.

D. Statistical Evaluation of PGP Measures of Productivity

We compare OPS and LWPGP on the basis of their ability to explain winning percentage across teams. The first three columns of Table 5 report coefficients from an OLS regression of team winning percentage on the two productivity measures. The explanatory power of the model using LWPGP dominates that of OPS by a comfortable margin (R^2 of .4640 vs. .4013). The third column of Table 5 reports coefficient estimates and test statistics from a regression including both measures. Both measures survive tests of the hypothesis that they add no information in the presence of the other, although for LWPGP the rejection of the null is much more decisive (a p-value of .0005 vs. .016), and the coefficient estimate for OPS has the "wrong" sign.

To incorporate the attributes of pitching and defensive ability, we construct measures of OPS and PGP achieved by a team's opponents throughout the season -- OPS against and LWPGP against. Columns 4 through 6 of Table 5 include these measures in the Win % regressions. The results yield a sharper contrast than those for offensive productivity measures alone. Again, the PGP-based model has higher explanatory power (R^2 of .8926 vs. .8682). In this case however, the increment in R2 from adding OPS to the PGP measures is trivial, and we fail to reject the hypothesis that OPS adds no

---

[6] Team OPS is calculated in the same way (as if all players on a team were the same batter). Note that OPS outperforms slugging percentage alone as a predictor of scoring.

[7] The normalization is based on coefficients from the OLS regression: OPS = .791 + .0055LWPGP (R2 = .985).

information to PGP (p-value of .699). On the other hand, LWPGP does add information to the regression with only OPS. Clearly, these tests confirm that the PGP measures are essential to predicting wining percentage of teams. In their presence, OPS adds little. By this standard, the linear weights-based PGP statistics are superior measures of productivity.

The final column of Table 5 breaks down OPS into components, on-base percentage (OBP) and slugging percentage (Slug Pct). The model is used to assess a claim made in *Moneyball*, that OBP is, contrary to decades of conventional wisdom, more important than slugging percentage on a point-for-point basis. The coefficients in this regression are consistent with this claim: those for OBP are 50 to 100% greater than the coefficients for Slug Pct. Nevertheless, breaking down OPS into its constituent components and allowing for separate coefficients still results in a model that is inferior to the model using LWPGPs.

III. Measurement of Clutch Effects & Other Noisy Productivity Measures

We have shown that PGPs are superior to other measures as predictors of team winning percentage. We now calculate individual player contributions in the same way, by assigning the impact of each play to the appropriate player and summing over all plate appearances (we ignore base running). Unadjusted PGP measures for each player sum the exact change in the probability of winning. The linear weights-based PGP measure (LWPGP) uses the average impact of each event. This measure ignores situation-specific game conditions, and can be used to help define "clutch" situations in various ways.

The most immediate way to define players who have performed in the clutch (and those who have not) is to simply difference their PGP and LWPGP measures. Those players who performed well in clutch situations will have a positive differential between their PGP and LWPGP. By contrast, players who "choked" in the clutch will have a negative differential between their PGP and LWPGP. Hence, method 1 defines clutch performance by the difference in probability impact as weighted by actual and average situations. For purposes of comparison across methods, we present PGP and LWPGP measures per plate appearance.

Since unadjusted PGP totals can vary a great deal based on a small sample of plays, we consider two additional methods of identifying clutch effects. Method 2 uses the baseball statistician's term of "late and close:" games in the 7^{th} inning or later in which the difference in score is two runs or less. Situations near the end of a close game are more critical than others. Hence, those whose LWPGP totals per plate appearance are high in late and close situations relative to other situations can be regarded as clutch performers.

Method 3 uses our own probability metric to define key situations. A key situation is one where the probability impact of a play is more than twice as high as normal. This identifies 10.9% of plate appearances as key situations. Hence, our third measure is the difference in LWPGP per plate appearance in key situations and LWPGP per plate appearance in normal situations.

Finally, we consider a negative method -- a means of identifying players whose apparent productivity is higher in relatively meaningless situations. We define these situations to be when the probability impact of a play is less than ¼ that of a play in a normal situation. This identifies 16.0% of plate appearances as relatively meaningless. Players with a positive differential in LWPGP per plate appearance in meaningless vs. normal situations are potential "cripple-killers."

It is obvious that in any given season, some players will appear at the top of each list even if clutch or cripple-killing is completely random. There must be a systematic component for either phenomenon to be meaningful. Hence, the existence of clutch effects can be proven if players who tend to be "clutch" in one season perform similarly in other seasons. At the present time, we are limited to two seasons of data, 1999 and 2000. Are the clutch performers from 1999 better than average in 2000? To answer this question, we partition each season's clutch measure into quartiles for all full time players.[8] If clutch performance is systematic, players in the top (bottom) quartile in one season are more likely to be in the top quartile in other seasons than players who were not. If the phenomenon is random, players will be distributed randomly across quartiles in successive seasons. The null hypothesis of randomness can thus be tested using simple cross-tabulations and a chi-squared test of independence.

These tabulations and associated test statistics are presented in Table 6. There is no evidence to support the hypothesis that clutch performers are persistent across seasons. In the case of cripple-killers, the test-statistic rejects randomness, but this is because there are too *few* players who repeat their performance in the top and bottom quartiles.[9]

We believe that using PGP based measures allow for the possibility of clutch performance to reveal itself in ways that traditional measures obscure. Nevertheless, these results confirm the results of earlier studies by Cramer (1977) and Brooks (1989). By our definition, clutch performers in 1999 are no more likely to be clutch performers in 2000 than any other player. While we would like to have more data to evaluate this more thoroughly, to date we know of no study which demonstrates that any systematic or persistent nature exists in clutch performance.

IV. Market Valuation of Noisy Production

[8] We define a full time player in this context as one with 50 or more plate appearances in clutch situations in each season.

[9] There is one notable entry in the cripple-killer table that gives one pause for thought: Sammy Sosa of the Cubs is identified as a cripple-killer in both 1999 and 2000. It is possible that his differential LWPGP in meaningless situations stems from a) the Cubs being behind with the game out of reach, and b) the opponents being willing to pitch to this popular home run hitter in these situations (and not others).

We now turn to the question of whether clutch performance is valued in the labor market. Although there is no evidence that clutch performance is persistent, there are reasons to believe that belief in the existence of clutch performers may impact the labor market. First, belief in clutch performers is likely to be quite strong. Although our statistical efforts have failed to document its presence in these data, these methods may lack statistical power against the null hypothesis of random differences in performance. Indeed, these results convince us not that clutch performance is a myth, but rather (more modestly) that it is very difficult to document with standard statistical methods even it is for real.[10]

Second, a player having come through in the clutch has demonstrated that he is an unlikely candidate for being a "choker," i.e. one who "freezes" or "tightens up" excessively in critical situations. This attribute may be valued in the labor market, perhaps properly so. Our measure of clutch performance -- a positive differential productivity in tight situations -- may thus be associated with player salary.

To detect remunerative effects of clutch performance we narrow our focus in two ways. First, we attempt to explain player salary using only recent performance measures, including clutch performance. Second, we restrict attention to players whose contracts changed by an appreciable amount, $150,000 or more.[11] We have two years of performance data, from 1999 and 2000. We use performance statistics from these years as explanatory variables for salary in the following year. We pool data from both years, but since market conditions change, we include a dummy variable for the 2001 contract year in all regressions. In addition, the market for free agents is relatively competitive, whereas that for younger players is much less so, relying on arbitration to limit monopsony power of owners. We include an indicator variable "arbitration" for those players not yet eligible for free agency.

The empirical model to be estimated then is

$$\begin{aligned}\ln(\text{Salary}) = \text{Constant} + b_0{*}2001\ \text{Dummy} + b_1{*}\text{Arbitration} +\\ b_2{*}\text{Clutch} + b_3{*}\text{Non-Clutch Performance} + \varepsilon\end{aligned}$$

All performance measures are calculated on a per plate appearance basis. Clutch is the difference between PGP and LWPGP per plate appearance. Our basic measure of Non-Clutch performance is LWPGP per plate appearance.

We also include models using more traditional performance measures -- OBP, Slug Pct, and RBI per plate appearance. The labor market's valuation of OBP and Slug Pct can be assessed relative to their impact on game outcomes in Table 5. RBIs, like

---

[10] This is similar to the problem of detecting the "hot hand," streaky performance relative to a baseline measure.

[11] We also require that a player have 502 or more plate appearances. The $150,000 minimum increment may be too low, as many players have escalating contracts that are predetermined several years in advance. But this works against finding any impact of clutch, or any other contemporaneous performance measure on salary.

Clutch, are dependent on many random factors outside the player's control: men on base, batting order position as determined by the manager, etc. The sabermetricians studied in *Moneyball* claim that RBIs are an uninformative statistic relative to Slug Pct and OBP. RBIs are determined in part by systematic factors under a player's control -- Slug Pct and OBP -- and in part by factors out of his control, as mentioned above. RBIs, like clutch, should be unrelated to salary in a market which values player contributions purely on the skills which increase a team's chances to win.

Table 7 presents coefficient estimates and summary statistics from the player salary regression. All models are similar in their ability to explain variation in player salaries (R2 varies from .494 to .506). Player performance as measured by LWPGP has a large and statistically significant impact on salary. Clutch PGP's are associated with a coefficient estimate of about half the size of LWPGP and twice its standard error. Clutch performance is positively valued, but of borderline significance in the basic model (p-value of .073).

Since LWPGP takes account of runs batted in through the impact of each plate appearance on the probability of winning, adding RBIs per plate appearance to LWPGP should, in a manner similar to clutch, merely add noise to the regression.[12] RBIs, though random, are less noisy than Clutch. In addition, RBIs are a traditional statistic that receives substantial attention. For these reasons we expect RBIs may contribute to the salary regression in a manner similar to Clutch, but perhaps with a lower standard error. The explanatory power of the regression is little changed by substituting RBIs for Clutch (R2 rises from .494 to .499). RBIs are positively valued over and above the information already inherent in PGPs, and the coefficient estimate is statistically significant (the p-value is .02). Including both Clutch and RBIs simultaneously in the regression yields results similar to those discussed above.

Finally, we include a model based primarily on traditional performance statistics: OBP and Slug Pct. What is notable in this regression is that Slug Pct is more highly valued in the labor market than OBP (by about 75%). This is a sharp reversal of their relative impact on team winning percentage. In addition, Clutch effects are again positively valued and of borderline significance (the p-value is .077). Not so OBP, to some the most important performance statistic in baseball (its p-value is .244). Labor market valuation of these skills is not commensurate with their impact on winning baseball games. Overall, these results provide strong hints, though not yet definitive evidence, that performance attributes which are not predictably related to the probability of winning baseball games are positively valued in the labor market.

V. Conclusion

Although much discussed by fans of the game, there has been little systematic study of the phenomenon of clutch performance in the academic literature. Our

---

[12] In future versions we will examine this claim of the sabremetricians. If they are correct (and we have no reason to doubt them), RBI that cannot be predicted on the basis of OBP and Slug Pct and position in the batting order are random between seasons.

procedures employed in this paper improve on those used by baseball statisticians. Specifically, we take account of the fact that clutch feats may lie in the ability to slap a two run single or hit a sacrifice fly in crucial situations. The impact of such plays on baseball games are often significantly greater than that of a majestic home run with no men on base, but are ignored or obscured by traditional statistics. Our measures of productivity are shown to be exceptional predictors of game outcomes. Despite this advance, like others before us we find no evidence that clutch performances are persistent across individuals.

There is reason to believe, however, that belief in clutch performance may impact the labor market. We report suggestive evidence of such an impact. In addition, we find persuasive evidence that the returns to skills in the labor market are not aligned with their impact on winning baseball games. We speculate that the labor market may be following fan preferences -- and not strictly a production function based on winning -- in salary determination. This is a topic we are pursuing in ongoing research.

Table 1: Probability of Runs Scored in an Inning, by Basecode and Outs

base code	out	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

Data Source: Stats Inc. Play-by-Play data for 1999.

Basecodes: 1 = runner on 1st; 2 = runner on 2nd; 3 = runner on 3rd; 4 = runners on 1st and 2nd; 5 = runners on 1st and 3rd; 6 = runners on 2nd and 3rd; 7 = bases loaded.

Note: 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 2: Impact of Changes in States on the Probability of Winning

Inning	Home/ Visitor	Run Difference	Basecode Before	Outs Before	Probability of Winning	Event
A. Early vs. Late Plays						
Value of Early Plays						
1	Visitor	0	0	0	0.500	
1	Visitor	0	1	0	0.534	Single/1B Error
1	Visitor	0	2	0	0.552	Double
1	Visitor	0	3	0	0.584	Triple
1	Visitor	1	0	0	0.590	Homer
Value of Late Plays						
9	Visitor	0	0	0	0.500	
9	Visitor	0	1	0	0.579	Single/1B Error
9	Visitor	0	2	0	0.660	Double
9	Visitor	0	3	0	0.770	Triple
9	Visitor	1	0	0	0.821	Homer
B. Changes in the Value of Leads						
Increasing Value of a Small Lead						
3	Visitor	1	0	0	0.604	
6	Visitor	1	0	0	0.648	
9	Visitor	1	0	0	0.821	
Increasing Value of a Large Lead						
3	Visitor	3	0	0	0.779	
6	Visitor	3	0	0	0.848	
9	Visitor	3	0	0	0.960	
C. The Bottom Half vs. the Top Half						
1	Home	0	0	0	0.551	
9	Home	0	0	0	0.653	

Note: The probability value applies to the team given by the team indicator.

Table 3: Average Effect of Selected Events Upon Probability of Winning

Event	Frequency	Mean Change in P(win)	Std. Error	Normalized Palmer Wt.	Norm Palmer/ P(win)
Walk	17,028	0.0281	0.0002	.0300	1.067
Hit by pitch	1,572	0.0284	0.0006	.0300	1.056
Single	29,686	0.0418	0.0003	.0418	1.000
Double	8,902	0.0646	0.0007	.0727	1.125
Triple	952	0.0948	0.0026	.0927	0.977
Home run	5,693	0.1217	0.0013	.1272	1.045
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		
1B error	1,549	0.0375	0.0010		
2B error	247	0.0595	0.0046		
3B error	25	0.0920	0.0222		
Stolen base	2,531	0.0137	0.0003	.0273	1.990
Caught stealing	1,308	-0.0428	0.0008	-.0545	1.274

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. For purposes of comparison, Palmer's weights (based on the impact of each event on scoring runs) are normalized by the impact of singles. Singles and triples appear to be relatively undervalued by Palmer's weights.

Table 4: Team Offensive PGP Totals and OPS for 1999

Rank	Team	PGP	LWPGP	OPS	Adj 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

Note: Teams are ranked by LWPGP. Adj PGP is normalized to the scale of OPS (using OLS regression coefficients).

Table 5 Linear Weights-Based PGPs Predict Win % Better Than OPS Statistics

	Model						
	1	2	3	4	5	6	7
Constant	0.516 (0.007)	-0.494 (0.159)	2.697 (0.879)	0.500 (0.003)	.483 (0.102)	0.801 (0.681)	.400 (0.124)
LWPGP	0.008 (0.001)		0.023 (0.006)	0.0071 (0.0005)		0.0072 (0.0034)	
LWPGP Against				0.0069 (0.0005)		0.0048 (0.0026)	
OPS		1.274 (0.204)	-2.754 (1.110)		1.325 (0.097)	-0.0029 (0.5964)	
OPS Against					(-1.304) (0.092)	-0.3766 (0.4649)	
OBP							2.0060 (0.3849)
OBP Against							-1.5539 (0.3768)
SlugPct							0.9898 (0.2064)
SlugPct Against							-1.1179 (0.2227)
N	60	60	60	60	60	60	60
R^2	.4640	.4013	.5162	.8926	.8682	.8939	.8768
F-test of H0: OPS coefficients = 0 p-value			6.153 0.016			0.360 0.699	
F-test of H0: PGP coefficients = 0 p-value			13.536 0.0005			6.667 0.0025	

Note: Standard errors are in parentheses.

Table 6: The Distribution of Clutch Performers Across Seasons

A. PGP - LWPGP per Plate Appearance					
		2000 Season			
		Top Quartile	Mid Quartiles	Bottom Quartile	Sums
	Top Quartile	10	25	16	51
1999 Season	Middle Quartiles	27	56	29	112
	Bottom Quartile	17	31	7	55
	Sums	54	112	52	218
	Test Statistic	6.03			
	P-Value	.197			
B. Difference in LWPGP per Plate Appearance: Late and Close vs Normal					
		2000 Season			
		Top Quartile	Mid Quartiles	Bottom Quartile	Sums
	Top Quartile	13	28	13	54
1999 Season	Middle Quartiles	29	54	27	110
	Bottom Quartile	12	28	14	54
	Sums	54	110	54	218
	Test Statistic	0.39			
	P-Value	.983			
C. Difference in LWPGP per Plate Appearance: Key vs Normal Situations					
		2000 Season			
		Top Quartile	Mid Quartiles	Bottom Quartile	Sums
	Top Quartile	6	16	7	29
1999 Season	Middle Quartiles	15	29	14	58
	Bottom Quartile	8	13	8	29
	Sums	29	58	29	218
	Test Statistic	0.72			
	P-Value	.948			
D. Difference in LWPGP per Plate Appearance: Blowout vs Normal Situations					
		2000 Season			
		Top Quartile	Mid Quartiles	Bottom Quartile	Sums
	Top Quartile	10	25	15	50
1999 Season	Middle Quartiles	19	54	28	101
	Bottom Quartile	21	22	7	50
	Sums	50	101	50	201
	Test Statistic	11.70			
	P-Value	.020			

Note: If clutch performance was systematic, the top quartile of performers in 1999 would be concentrated in the top quartile in 2000 also. In sections A-C however, the allocation is well described by a distribution in which players from each quartile from 1999 are equally like equally to be land in any quartile in 2000. In section D, the null hypothesis of an equally likely allocation is rejected, but this is due to apparent reversals in performance, rather than persistence.

Table 7: Is There a Labor Market Impact of Clutch Performance?

Dependent Variable: Ln(Salary)

Variable	Model 1	Model 2	Model 3	Model 4
Constant	15.170 (0.072)	15.126 (.075)	15.126 (.074)	15.128 (.074)
2001 Dummy	0.214 (0.072)	0.227 (.072)	0.219 (.072)	0.235 (.072)
Arbitration	-.866 (0.074)	-.857 (0.073)	-.839 (0.073)	-.862 (.073)
LWPGP	86.686 (12.045)	66.017 (15.272)	67.137 (15.246)	
Clutch	42.352 (23.521)		35.074 (23.604)	41.351 (23.300)
RBI		2.981 (1.293)	2.692 (1.303)	
OBP				1.607 (1.375)
SlugPct				2.930 (.646)
Observations	214	214	214	214
R^2	.494	.499	.504	.506

Performance variables are (a) measured per plate appearance and (b) relative to replacement level. See Table A.1 for replacement levels of various performance measures. The sample includes all players with (a) a minimum of 502 plate appearances each season and (b) an increase in salary of $150,000 or more

Table A.1: Replacement Level Statistics by Position

Position	Player	LWPGP/PA	OPS	OBP	SP	RBI/PA
Catcher	Henry Blanco Mike Matheny	-.0031	.6823	.3098	.3560	.0909
1B/DH	Rico Brogna Erik Karros	-.0010	.7886	.3402	.4292	.1254
2B	Lenny Harris Mark Grudzielanek	-.0017	.7263	.3263	.3657	.0781
3B	Travis Fryman Bill Muehler	-.0030	.7185	.3086	.3862	.1109
SS	Mike Mordecai Alex Cora	-.0031	.6600	.3071	.3425	.0880
LF	Gerald Williams B.J. Surhoff	.0002	.7860	.3421	.4483	.1187
CF	Marquis Grissom Marvin Benard	-.0017	.7413	.3227	.3844	.0830
RF	Bob Higginson Matt Stairs	-.0009	.7332	.3353	.4058	.1129

Each number in this table corresponds to the value of the respective statistic at the 20^{th} percentile of each distribution.

Note: The players named are full-time players with offensive statistics similar to those listed. The top player for each position is drawn from 1999 statistics, while the player underneath was selected based on 2000 statistics.

Appendix 2a: The value of starting pitching - probability based measures from the betting market

This appendix illustrates a method whereby information from the betting market can be used to measure the value of starting pitching. In the seminar, we will compare these measures to measures based on the methods used in the main body of the paper. The analysis pursued here is essentially a test of the "Bill James Conjecture." In his recent book *Win Shares*, James attempts to measure the number of wins accounted for by each player on a baseball team. To do this, he must take a stand on the relative importance of hitting, pitching, and fielding. His conjecture is that half the game is offense and half defense, with 2/3 of the defensive component attributable to pitching. Hence, pitching is 1/3 rd of the game.

In our test of the conjecture, we assign $1/3^{rd}$ of the probability change associated with each event to the pitcher (pgp_i). We sum these over the season, and divide by the number of starts to get a per game measure for each pitcher ($pgppg_i$), based on our methods and James' conjecture.

Implicit in the betting odds are the market's estimate of the impact of each pitcher on the probability of winning the game. We find a close – nearly 1 to 1 - relation between the market estimates and our pgp measures.

There are 2385 observations (games) in 1999, R-squared = 0.8631.
There are 2390 observations (games) in 2000, R-squared = 0.8191.
Each season is estimated separately.

$$Pr\_h = \beta_0 + \beta_1 * h\_ops + \beta_2 * a\_ops + \beta_3 * D_1 + \beta_5 * D_2 + \beta_5 * D_3 + \varepsilon$$

The dependent variable is the probability the home team wins each game, as calculated from the betting line and corrected for vigorish.

D1 is a vector of dummy variables for each home team.
D2 is a vector of dummy variables for each away team.
D3 is a vector of dummy variables for each starting pitcher.

The home team's OPS and away team's OPS are included as independent variables and are calculated with statistics from all games prior to the day of the game being examined. For games in the first 40 games of the season, h_ops and a_ops are averages of the current season's OPS and the previous season's OPS, with the weight on the current season being (game # / 40), and the weight on the latter is (1 – game # / 40). For games after each team's 39th of the year, all OPS data is from the current season.

A dummy variable for each home team and for each away team was included (controlling for defense, relief pitching, managerial talent, etc.

A pitcher is categorized as a starting pitcher if he started more than 10 games in either of the 1999 and 2000 seasons. A dummy variable for each starting pitcher was included in the regression, taking the value 1 if the pitcher was the starter at home, and taking the value -1 if the pitcher was the starter away. All pitchers who were not classified as a starting pitcher as described above are grouped together and given a dummy variable taking the value 1 at home, and -1 away.

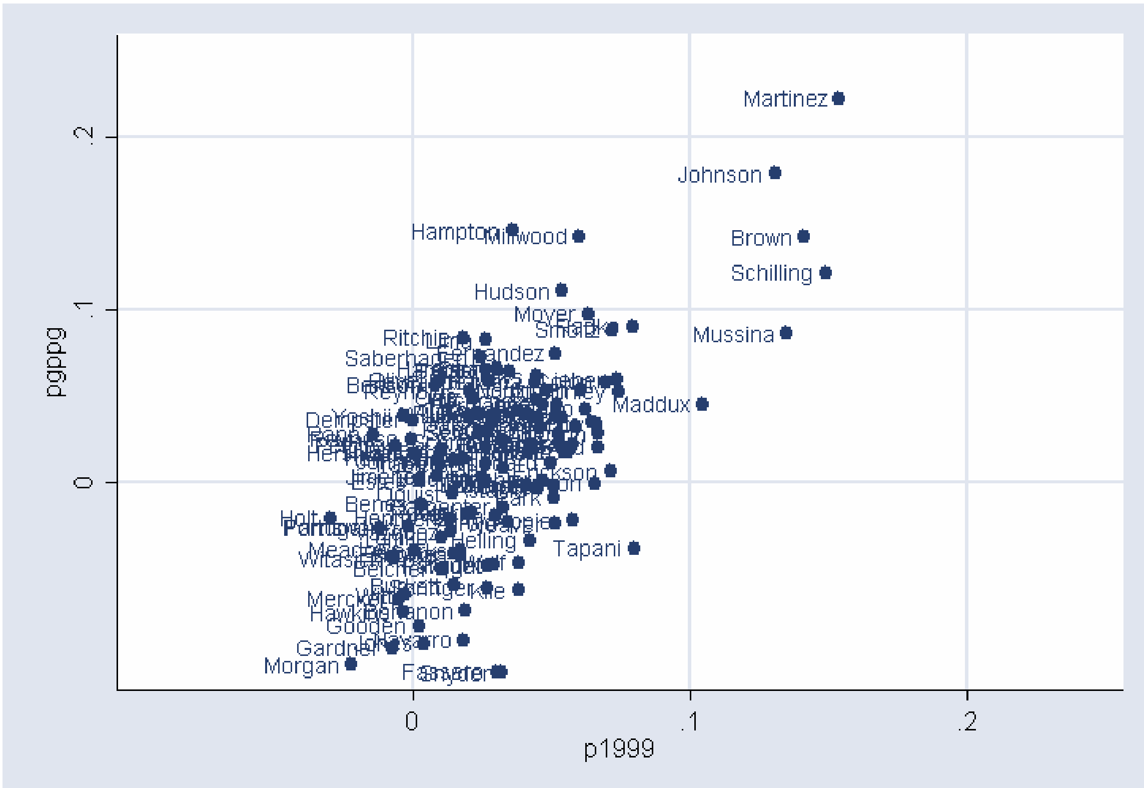
Coefficient estimates (other than the dummies) from the 1999 season are given below:

Variable	Coef.	Std. Err.	t
_cons	0.504	0.020	25.8
homedum9999	0.004	0.002	2.55
homeopsdb	0.076	0.016	4.74
awayopsdb	-0.055	0.017	-3.2
Adj R-squared	0.851		
Root MSE	0.036		

Below are estimates for pitchers that are the largest in magnitude. These coefficients can be interpreted as the market's estimate of the impact of that pitcher on his team's probability of winning the game (over and above the impact of a replacement pitcher).

1999		2000	
P.Martinez	0.153852	P.Martinez	0.226191
C.Schilling	0.148916	R.Johnson	0.170888
K.Brown	0.141394	K.Brown	0.144966
M.Mussina	0.135116	G.Maddux	0.111094
R.Johnson	0.130619	M.Mussina	0.109761
G.Maddux	0.104611	A.Leiter	0.089289
K.Tapani	0.079895	C.Schilling	0.088918
B.Radke	0.079373	D.Wells	0.086299
C.Finley	0.074121	M.Hampton	0.075809
J.Lieber	0.073567	B.Radke	0.070979
----------	----------	----------	----------
M.Yoshii	-0.00298	B.Rekar	-0.00285
L.Hawkins	-0.00319	A.Reynoso	-0.00566
K.Mercker	-0.00469	B.Meadows	-0.00783
A.Pettitte	-0.00582	M.Mulder	-0.01159
J.Witasick	-0.00685	D.Mlicki	-0.01539
M.Gardner	-0.00719	B.Penny	-0.0161
M.Portugal	-0.01183	J.Sanchez	-0.01697
P.Rapp	-0.01408	P.Schourek	-0.0202
M.Morgan	-0.02192	J.Witasick	-0.02127
C.Holt	-0.02949	K.Wells	-0.02413

There is a close correspondence between these estimates and the PGP measures for pitchers. The figure below maps the estimates on the horizontal axis, and the James-adjusted pgp per game measurement on the vertical.



Below are coefficients from a regression of pgppg on the market's probability estimates for each pitcher (referred to by p1999 in the graph, Pitcher_p below).

Variable	Coef.	Std. Err.	t-ratio
_cons	-0.023	0.006	3.77
Pitcher_p	1.069	0.127	8.41

Number of obs = 127
R-squared = 0.3616
Adj R-squared = 0.3565
Root MSE = .04627

test (Pitcher_p=1)
F(1, 125) = 0.29
Prob > F = 0.5905

test (_cons=0) (Pitcher_p=1)
F(2, 125) = 13.02
Prob > F = 0.0000

James' conjecture thus gets some, but not complete support from the betting market data. If the market is correct, the negative and significant constant implies that James' conjecture slightly undervalues pitching. This is preliminary however, and needs to be assessed with data from additional seasons.

References

Albert, Jim and Bennett, Jay. *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*, Springer Verlag, 2001: New York.

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.

Brooks, Harold. "The Statistical Mirage of Clutch Hitting," Baseball Research Journal, Sixth Annual Historical and Statistical Review of the Society for American Baseball Research, 1989, pp. 63-66.

Brown, William O. and Sauer, Raymond D. "Does the Basketball Market Believe in the Hot Hand: Comment," *American Economic Review*, December 1993.

Burgess, Paul L., Marburger, Daniel R., and Scoggins, John F. "Do Baseball Arbitrators Simply Flip a Coin?" In *Baseball Economics*, J. Fizel et al eds., Praeger, 1996: Westport, CT.

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

Cramer, Richard C. "Do Clutch Hitters Exist?" Baseball Research Journal, Sixth Annual Historical and Statistical Review of the Society for American Baseball Research, 1977.

"Fizel, John P. "Is There Bias in Major League Baseball Arbitration?" In *Baseball Economics*, J. Fizel et al eds., Praeger, 1996: Westport, CT.

Fort, Rodney. "Pay and Performance: Is the Field of Dreams Barren? In *Diamonds are Forever*, P. Sommers ed., Brookings, 1992: Washington D.C.

Gius, Mark Paul and Hylan, Timothy R. "An Interperiod Analysis of the salary Impact of Structural Changes in Major League Baseball: Evidence from Panel Data," In *Baseball Economics*, J. Fizel et al eds., Praeger, 1996: Westport, CT.

Hakes, Jahn K., and Sauer, Raymond D. "Effects and Valuation of Fielding in Major League Baseball: A Play-by-Play Analysis," manuscript, July 2002 Clemson University.

James, Bill and Henzler, Jim. *Win Shares.* Stats Inc. Publishing, 2002: Morton Grove, IL.

Kahn, Lawrence M. "Management Quality, Team Success, and Individual Player Performance in Major League Baseball," *Industrial and Labor Relations Review*, 1993a, 46, 531-547.

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 1993b, 75, 157-164.

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

Marburger, Daniel R. "A Comparison of Salary Determination in the Free Agent and Salary Arbitration Markets," In *Baseball Economics*, J. Fizel et al eds., Praeger, 1996: Westport, CT.

McDonald, Don N. and Reynolds, Morgan O. "Are Baseball Players Paid their Marginal Products," *Managerial and Decision Economics*, September/October 1994, 15, 443:457.

McCormick, Robert E. and Tollison, Robert D. "Crime on the Court," Journal of Political Economy, v. 92, no. 2, 1984: 223-35.

Neyer, Rob. "Clutch Hitting," at Diamond Mind Baseball, http://www.diamond-mind.com/articles/neyerclutch.htm, visited June 9, 2003 (originally published Dec. 17, 1999 on ESPN.com).

Romer, David. It's Fourth Down and What Does the Bellman Equation Say? A Dynamic-Programming Analysis of Football Strategy, NBER Working Paper 9024, February 2003.

Ruggiero, John, Hadley, Lawrence, and Gustafson, Elizabeth. "Technical Efficiency in Major League baseball," In *Baseball Economics*, J. Fizel et al eds., Praeger, 1996: Westport, CT.

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

Sommers, Paul M. and Quinton, Noel. ""Pay and Performance in Major League Baseball: The Case of the First Family of Free Agents," *Journal of Human Resources*, Summer 1982, 17: 426-36.

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