Salaries, Performance, And Owners' Goals In Major League Baseball: A View Through Data

Author(s): M. R. Yilmaz and Sangit Chatterjee

Source: *Journal of Managerial Issues*, Summer 2003, Vol. 15, No. 2 (Summer 2003), pp. 243-255

Published by: Pittsburg State University

Stable URL: https://www.jstor.org/stable/40604428

JOURNAL OF MANAGERIAL ISSUES
Vol. XV Number 2 Summer 2003: 243-255

Salaries, Performance, And Owners' Goals In Major League Baseball: A View Through Data

M. R. Yilmaz
Professor of Business Administration
Northeastern University

Sangit Chatterjee
Professor of Business Administration
Northeastern University

It is no secret that economics has become a prominent part of the "business" of sports in recent decades, and not just at the professional level. Even the Olympic Games, the ultimate competitive forum for all kinds of sports, could not sustain the ideal of pure competition among amateur athletes. Today, player salaries, team revenues, broadcasting and media contracts, merchandise marketing, etc. are as much a part of sports as the games themselves. This is particularly true of Major League Baseball, the oldest of professional sports leagues, and the subject of our study. Several recent books have appeared in the literature to explore different aspects of economic factors on baseball. Among these, Leeds and von Allmen (2002) examine the economics of professional and college sports industry, including baseball. Abrams (2000) focuses on free agency and salary arbitration in baseball from a legal perspective. Morgan and Lally (1999) give a rather pessimistic account of the labor problems and future prospects of baseball, and Kelley (1996) takes a critical look at the bonus contracts that used to be given to unproven baseball prospects. Fizel *et al.* (1996) present a survey of research in baseball economics, and Scully (1995) examines the market structure of sports from a purely economic perspective. Sommers (1992) provides a collection of articles and opinions on the business of baseball.

The quantitative literature on baseball has also been developing rapidly over the years. Most of the current literature deals with baseball as a market phenomenon in which owners are the firms in the business of entertainment while the players are the ones supplying the input to the production process. Examples of such articles are provided by Ferguson *et al.* (2000) and Richards and Guell (1998). Numerous articles deal with the effects of free agency and arbitration on salary structures (e.g., Miller, 2000; Bodvarsson and King, 1998; Fizel, 1996). The article by Marburger (1994) examines the effect of bargaining power on salaries, and Krautman (1999) discusses the difficulties in using salary as an estimate of a player's marginal revenue product. There are also articles dealing with race as a factor in determining salaries (e.g., Marburger, 1996), or even the market for baseball cards (McGarrity *et al.*, 1999).

There have also been several studies dealing with fan attendance in baseball. Kahane and Shmanske (1997) consider the relationship between team roster turnover and attendance. Knowles *et al.* (1992) study the relationship of demand for tickets to the uncertainty of the outcome of the game, and Domazlicky and Kerr (1990) consider the relationship between baseball attendance and the designated hitter. Whitney (1988) investigates fan interest (attendance) and team performance (winning games versus winning championships). At the aggregate level, Schmidt and Berri (2001) found that there is a significant relationship between competitive balance and attendance in Major League Baseball. The study by Horowitz and Zappe (1998) examines the determinants of baseball veterans' end-of-career salaries as a function of past performance and other criteria, and it is akin to the spirit of our study, but with a more specialized focus.

In this study, we consider player salaries and performance at the individual as well as the team level using recent data. Our purpose is to see if individual and team performance measures that are associated with player compensation are different or similar to those associated with team owners' objectives. Research related to this question has not been reported in the literature. Thus, our main contribution is the examination of game performance characteristics that is of economic interest to both the players and team owners. Although the scope of this study is limited to players other than pitchers, we found some interesting results in terms of the underlying variables of performance from the perspectives of players and the owners.

Most firms view compensation as a means of motivating as well as retaining employees. If compensation is closely tied to key performance measures, then it is reasonable to expect that employees work to maximize those measures in order to maximize their compensation. Further, if performance measures are chosen which maximize the firm's performance and are most likely to help the firm attain its goals, then the objectives of employees and the firm are aligned. In many organizations, achieving "performance-based pay" may be difficult to implement, and in others, selecting the correct measures may be a problem. Sales representatives are often compensated in accordance with the sales volume or revenue they generate. In many cases, sales are made at a loss near the end of an accounting period, and representatives can "make their numbers." In such cases, the employees have responded predictably to their compensation plans, but not necessarily in the best interest of the firm.

The measurement of performance is less of a problem in professional sports than in most other industries. In particular, Major League Baseball (MLB) offers a wealth of measures pertaining to a player's performance on the playing field. In this study, we considered the relationship of individual performance to salaries and to the performance of teams. We examined the important characteristics of player performance as they may influence the compensation received by the player. We then examined team performance from the perspective of the owners. Here, we conjecture that owners are primarily interested in the winning performance of their teams as well as maximization of revenues. Clearly, these objectives are neither identical nor entirely independent; they overlap to some extent. Some owners may emphasize winning (many people feel that George Steinbrenner of the New York Yankees is such an owner), while others may be more concerned with financial considerations. In any case, a relevant question is whether there is some alignment of important performance characteristics between the two parties—the players and the owners. From an organizational perspective, such an alignment is desirable because it enhances the likelihood of balance within the organization as well as the prospects for long-run success.

In the next section, we describe the performance measures and data used in this study. Then we present the analyses and results from the perspectives of the players and the owners. The last section includes a summary and concluding remarks about our study and the findings.

DATA

Since baseball games are decided by the number of runs scored, we used player performance measures that lead to run production. These measures are batting average (AVG), number of at bats (AB), runs scored (R), hits (H), home runs (HR), runs batted in (RBI), and bases on balls (BB). It is worth emphasizing that BB is the number of walks drawn by a hitter, not the number of walks issued by a pitcher. Data on these measures were collected for all MLB players in both leagues in the 1999 regular season, excluding playoffs. Starting salaries for the 2000 season were also collected. Players with data from only 1999 or only 2000 were excluded. Also excluded were players with fewer than 100 at bats in 1999, which is under 1 at bat per game, or just over 1 at bat every other game. Thus, occasional or temporary players, specialists with highly limited roles (e.g., a rarely used pinch hitter or runner), or players who were injured most of the season were excluded from the study.

Two clarifications should be made at this point. First, our use of player performance data in 1999 and salary data in 2000 reflects the view that a player's performance in a given period is a leading indicator of his salary in future periods. While the use of data from a single year is an obvious limitation in this regard, we felt that using lagged performance data would be an improvement over non-lagged data. Second, using 100 at bats in a season as the cutoff value to exclude some players could be questioned as being rather arbitrary. While this is true to some extent, we believe that some cutoff value should be used if the findings are to apply to the population of regular players, as opposed to those who are temporary, rarely-used, or have long-term injuries. The specific value itself is difficult to decide without controversy, but the following observation will be noted. When we tried different cutoff values between 75 and 150 at bats, there were no significant changes in the results obtained with 100 at bats. For example, at 75 at bats, the sample size increased slightly from 362 to 371, but there were no significant differences in the summary characteristics, given in Table 1 below. At 150 at bats, the sample size decreased by 32 players, but again, there were no significant changes in the reported summaries.

For the 30 teams in the National and American leagues, the mean, minimum and maximum of the above measures for each team were used as possible explanatory variables for team performance. Overall team performance and success were represented in terms of two response variables, namely the *number of games won* and *total home attendance* for the year 1999. Since team revenue figures are generally not available, total home attendance was used as a proxy for revenues or financial success. All data we used were available on the Internet at *www.espn.com* and *www.stats.com*.

The distribution of player salaries is shown in the top histogram in Figure I. The mean salary for all 362 players with at least 100 at bats was $2,491,000, and salaries varied from a minimum of $200,000 to a maximum of $12,868,000. As expected, the distribution is very highly skewed to the right. Consequently, several transformations were used to see if they would reduce or eliminate skewness. Though skewness could not be eliminated entirely, it was considerably reduced by the logarithmic transformation. The resulting histogram is shown at the bottom in Figure I using logarithms with base 10. It is seen that the left side of the histogram continues to exhibit skewness, whereas the right side is relatively symmetric and bell-shaped.

Based on this last observation, data were divided into two parts, consisting of players whose salaries were below $1 Million in one group, and at least $1 Million in the other. This is shown in the bottom graph in Figure I, with $1 Million positioned at 6 on the logarithmic scale. Means and standard deviations of individual performance measures in the two groups are shown in Table 1. It is seen that the higher-paid players have higher mean values on all measures. In particular, the means for at bats (AB) clearly show that higher-paid players (n = 200, mean AB = 453) also play much more regularly than lower-paid players (n = 162, mean AB = 296).

It should be noted that dividing the data in the manner described above is more than a matter of convenience. We believe that the bimodal distribution at the bottom of Figure I indicates the presence of two sub-populations comprising the population of all players. The higher-paid group consists of regular players who play in most games whereas the lower-paid group mostly includes occasional, temporary, injured, or role players who do not play on a regular basis. This is also supported by the vast difference in at bats statistics given above. Further evidence is provided by the following observation concerning the cutoff value we used for at bats. When we tried 75 AB as the cutoff value, the number of lower-paid players in the sample increased by 8, but the number of higher-paid players increased by only 1. When we tried 150 as the cutoff value, the number of lower-paid players decreased by 26, but the number of higher-paid players decreased by only 6, and several of these players had long-term injuries and other unusual circumstances. It seems clear that the

Figure I. The Distribution of Salary in 2000 (top) and Log Salary (bottom) for All Players With at Least 100 At Bats in 1999.

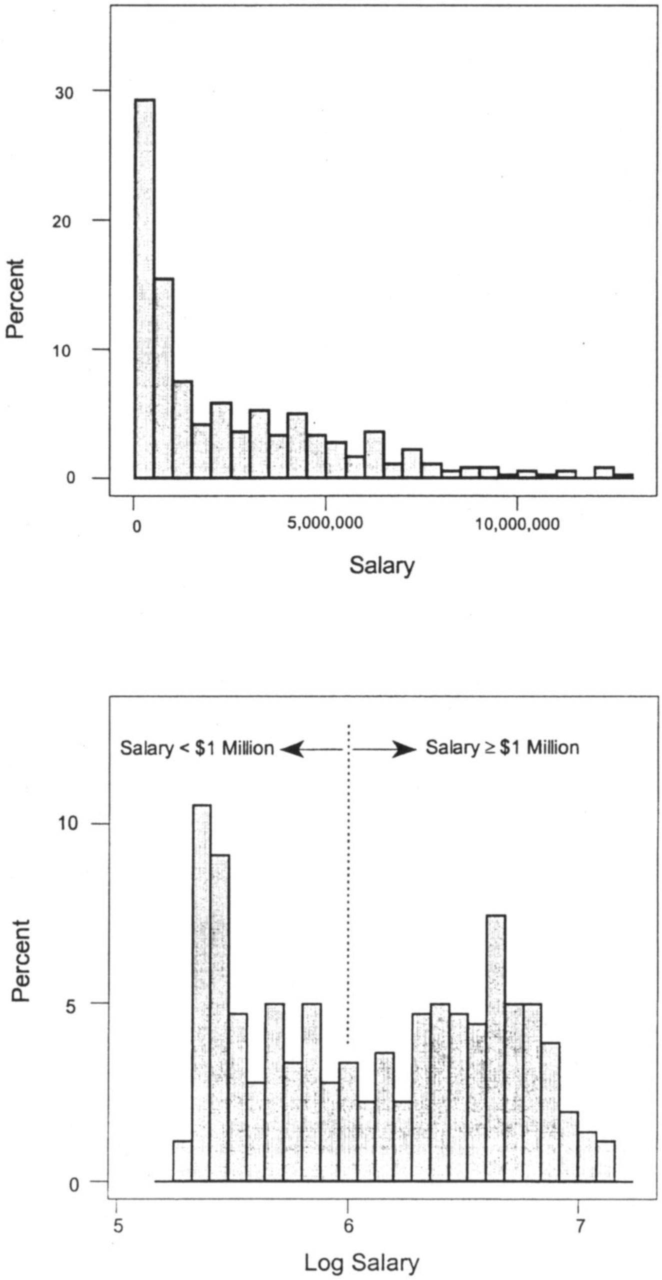


Table 1. Summary Statistics for Individual Performance Measures in 1999.

	Salaries below \$1 Million (n = 162)		***Salaries at least \$1 Million (n = 200)***	
Measure	***Mean***	***Std. Dev.***	***Mean***	***Std. Dev.***
Batting Average (AVG)	0.264	0.036	0.284	0.029
At Bats (AB)	296	139	453	144
Runs Scored (R)	40	23	74	31
Hits (H)	79	41	130	47
Home Runs (HR)	8	7	18	13
Runs Batted In (RBI)	39	24	71	34
Bases on Balls (BB)	29	17	54	28
Stolen Bases (SB)	8	3	10	3

Note: All values except AVG are rounded to the nearest integer.

lower-paid group has different characteristics than the higher-paid group.

An examination of scatterplots and Pearson correlation coefficients between all pairs of variables was performed, and the most important findings are shown in Table 2 (full correlation matrices are omitted for brevity). It is seen that all of the performance measures are significantly correlated with the logarithm of salary (and with salary) for higher-paid players, but none of the correlations are significant for lower-paid players. This is also visible in the scatterplots in Figure II. Thus, performance measures seem relevant for predicting salaries in the higher-paid group, but not in the lower-paid group. This observation prompted us to use the logarithm of salary for players in the higher-paid group as the response variable in the ensuing analyses. Since the data for lower-paid players were excluded from further analyses, our findings are not applicable to them.

ANALYSES AND RESULTS

The analyses were conducted separately from the perspectives of the players and the owners. The findings are reported accordingly.

The Players' Perspective

Regression models of Log Salary against various combinations of batting performance measures were examined. These analyses show that the best two-variable model for predict-

Table 2. Pearson Correlation Coefficients and Their p-values Between Individual Performance Measures and the Logarithm of Salary.

	Salaries below $1 Million (n = 162)		*Salaries at least $1 Million (n = 200)*	
Measure	***Correlation with Log Salary***	***Significance (p-value)***	***Correlation with Log Salary***	***Significance (p-value)***
Batting Average (AVG)	0.101	0.201	0.269	0.000
At Bats (AB)	0.016	0.843	0.366	0.000
Runs Scored (R)	0.033	0.678	0.469	0.000
Hits (H)	0.037	0.644	0.392	0.000
Home Runs (HR)	0.017	0.831	0.527	0.000
Runs Batted In (RBI)	0.035	0.656	0.540	0.000
Bases on Balls (BB)	0.109	0.166	0.452	0.000
Stolen Bases (SB)	0.031	0.623	0.232	0.000

Note: For players earning below $1 Million, none of the correlations are statistically significant ($p > 0.05$). This was also true for correlations with Salary without taking the logarithm.

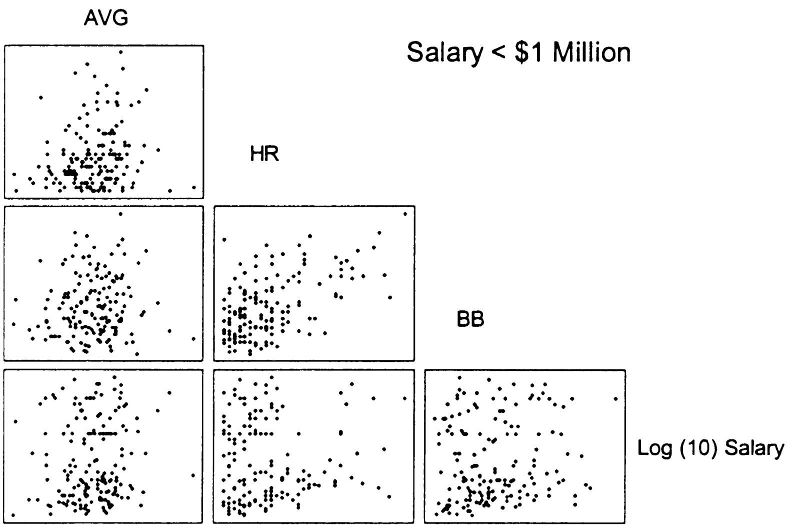
ing a player's salary includes RBI and BB as explanatory variables. This model explains 30.8% of the variability in Log Salary. Another two-variable model, which is almost as good, includes HR and BB and explains 30.7% of the variability in Log Salary. As expected, HR and RBI are highly correlated explanatory variables ($r = 0.903$). Players who hit a lot of home runs often have high RBIs, and vice versa. Other pairs of measures that exhibit multicollinearity were hits and at bats ($r = 0.964$), hits and runs scored ($r = 0.903$), and runs scored and at bats ($r = 0.889$).

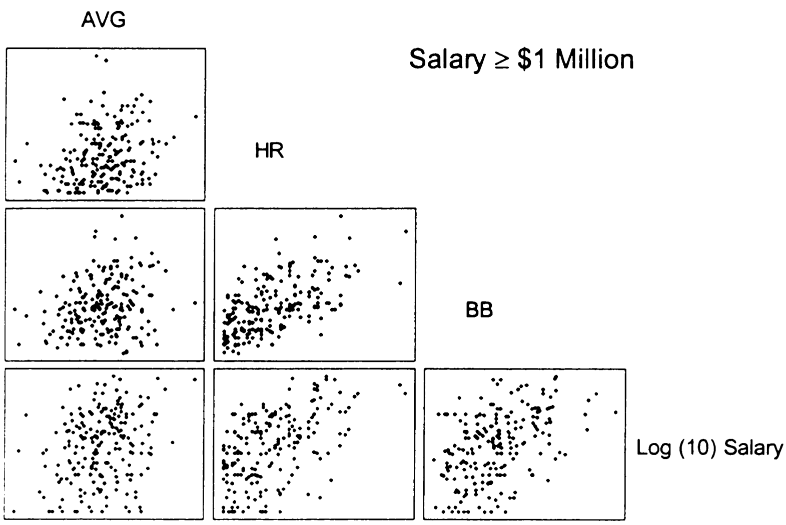
A slightly better three-variable model for Log Salary includes HR, BB and AVG. This model explains 32.2% of the variability in Log Salary. Although four-factor models were also considered, none of them offered a significantly better fit, and each of them included at least one variable that was not statistically significant. The best three-variable regression model is shown below (t-values are shown in parentheses underneath the coefficients). For this model, the coefficient of determination was R2 = 0.322 with the adjusted $R^2 = 0.312$.

$$\underset{}{\text{Log Salary}} = \underset{(36.98)}{5.94} + \underset{(2.06)}{1.19}\,\text{AVG} + \underset{(5.07)}{0.00814}\,\text{HR} + \underset{(2.71)}{0.00203}\,\text{BB}$$

So, if a player wants to maximize his salary, what skills does he need to develop? Clearly, an important skill is the ability to produce runs. One way to do this is to hit home runs, which always produces runs, and another is to get many hits, which often produces or leads to runs. Interestingly, it is also important to get on base by

Figure II. Scatterplots of Log Salary with AVG, HR, and BB Shown Separately for Players Making Less Than $1 Million and at Least $1 Million.





drawing walks. A player with disciplined batting skills, and not chasing bad pitches, will reach base much more often than a less disciplined hitter. Getting on base with a walk can also distract or upset the pitcher and thus help the hitters who follow. The significance of walks in terms of its affect on a player's salary is an interesting finding that has not been discussed elsewhere in the literature.

The preceding results shed some light on the skills owners are paying their players for, intentionally or not. But are these the skills owners want to maximize for their teams objectives?

The Owners' Perspective

If salary is intended to motivate players to maximize certain aspects of their performance, which aspects of team performance would motivate the owners? Some owners are true fans of baseball, and they are most interested in having a winning team. Others owners are investors looking to maximize the return on their investment or similar economic criteria. To examine these, we developed regression models for the number of games won during the regular season (Wins) as the measure of a team's on-field performance, and for total attendance in home games (Attendance) as a proxy measure for a team's financial success. As noted earlier, we used team counterparts of individual performance measures as possible explanatory variables. For each variable, data on the team average, minimum, and maximum were obtained. Before presenting the models, it is worthwhile to note that, although data are available for all 30 teams in MLB, it is appropriate to treat the data set as a sample rather than a population. Not only do the data belong to a single year, the data can also be considered to come from a conceptual population of all possible MLB teams.

Pearson correlations between various team performance measures and the response variables are shown in Table 3. For brevity, measures without a significant correlation with either response variable at the 0.10 level are not shown, although they are not omitted from the analysis. It is seen that variables such as Mean AVG, Mean H, Max H, Mean SB, Max SB, and any of the Minimum measures are not significantly correlated with winning or attendance. Also, high levels of multicollinearity between performance measures were not observed, with the exception of 0.821 correlation between Mean RBI and Mean HR. Somewhat surprisingly, the single best predictor of winning is Max BB, and the single best predictor of attendance is Mean HR. The latter finding is in agreement with the argument by Yilmaz *et al.* (2001) that home run hitting is a significant economic factor with significant impact on attendance.

A number of regression models were developed for Wins using up to seven variables. The most reasonable and parsimonious model among these included just two team performance measures, namely Mean RBI and Max BB. For this model, the coefficient of determination was $R^2 = 0.571$ (adjusted $R^2 = 0.539$). The model is given below with t-values shown in parentheses underneath the coefficients.

$$\text{Wins} = \underset{(2.52)}{25.6} + \underset{(2.86)}{0.433}\ \text{Mean RBI} + \underset{(4.42)}{0.307}\ \text{Max BB}$$

It is interesting to note that maximum number of walks emerges as a significant variable in explaining the

Table 3. Pearson Correlation Coefficients and Their p-values Between Team Performance Measures and Team Response Variables Wins and Attendance.

	Wins		***Attendance***	
Measure	***Correlation Coefficient***	***Significance (p-value)***	***Correlation Coefficient***	***Significance (p-value)***
Max AVG	0.318	0.087	0.466	0.009
Mean HR	0.492	0.006	0.575	0.001
Max HR	0.215	0.254	0.442	0.014
Mean RBI	0.511	0.004	0.422	0.020
Max RBI	0.423	0.020	0.511	0.004
Mean BB	0.378	0.039	0.369	0.045
Max BB	0.664	0.000	0.518	0.003
Mean SB	0.335	0.070	0.279	0.136
Max SB	0.357	0.053	0.229	0.223

Note: Measures without significant correlations are not shown.

number of games won. All else being equal, a team with a player who draws a lot of walks is significantly more likely to win games than a team that does not have such a player.

Finally, a number of regression models were developed and scrutinized for Attendance. The best of these models also includes three variables—Max AVG, Max HR, and Max BB—as depicted below:

$$\underset{(-3.36)}{\text{Attendance} = -6627022} + \underset{(3.65)}{21349300} \text{ Max AVG} + \underset{(2.23)}{24704} \text{ Max HR} + \underset{(2.35)}{11378} \text{ Max BB}$$

This model had a coefficient of determination of $R^2 = 0.552$ (adjusted $R^2 = 0.501$). Compared with the model for Wins, the reduction in R^2 in this model indicates that it is somewhat easier to predict the number of wins for a team than to predict attendance at its home games. Clearly, Wins and Attendance are correlated ($r = 0.487$), but fans attend the games for other reasons as well (e.g., to watch their favorite players). It is also interesting to observe that essentially the same measures that are significant in explaining the salaries of players (other than pitchers) are also significant in explaining attendance.

Does winning bring in revenue, which allows recruiting of better talent, which in turn enhances the chance of winning? Obviously there is a "chicken-and-egg" question here. Regression analysis is predicated on association rather than causality, and it is not possible to assert if one of these factors is the cause for the other. The truth is probably a mixture of both possibilities.

SUMMARY AND IMPLICATIONS

Perhaps the most interesting finding reported in our study is the fact that player skills associated with high salaries are the same skills that owners would want to pursue in order to enhance the likelihood of success of their teams in terms of winning or attendance. Home run hitting, ability to draw walks, and batting average emerge as the performance characteristics that are common to the interests of both sides. This alignment of skills required for individual as well as team success should be of interest to baseball researchers and fans alike.

Whether a team's owner is a business person interested in profits and returns on investment, or a baseball purist with the desire to win, the most desirable performance skills to encourage in the players are skillful batting leading to hits or walks, and the occasional slugging of home runs to score those runners. These findings are sometimes lost in the popular press coverage focused on player salaries and owner greed.

Our analysis has shortcomings on at least two basic counts. First, we used data for a single year in our analyses, and validation of the findings for other years would be desirable. Second, we used primarily offensive measures of player and team performance. The other side of this equation, of course, relates to the defensive performance, including pitching. Without taking pitching into account, the analysis remains incomplete. Nevertheless, analysis of players other than pitchers is still a useful pursuit in understanding the extent to which a common set of offensive performance variables may explain a portion of the variability in salaries (of non-pitchers), team winning, and attendance. Although bringing pitching into the equation is likely to improve explanatory power, it is also likely to remain less than complete, since factors other than player performance also influence the response variables (e.g., market size, television exposure). Further, performance variables for pitchers are quite different than the variables for non-pitchers, which detracts from the desirability of using a single equation for all players. A more reasonable alternative may be to use simultaneous structural equations, which would require a larger study than our data permit. In any case, it is clear that the salary equation we obtained cannot be used for pitchers.

In spite of these limitations, however, the findings seem new and interesting, not just to researchers but owners and players as well. These findings suggest that the two sides should understand each other very well in labor negotiations and collective bargaining, at least in terms of player salaries. In the current round of negotiations being conducted in early 2002 (after a delay of several months), salaries appear to be less of an issue than the proposed contraction in the number of MLB teams. Owners had proposed the elimination of two unprofitable teams (rumored to be Minnesota Twins and Montreal Expos), and the players' union had promptly opposed the idea as expected. This issue not withstanding, our findings suggest that the main determinants of individual player salaries are essentially the same performance characteristics that also enhance owners' goals. This should enhance common understanding and the chance of success in the negotiations for individual player salaries.[1]

References

Abrams, R. I. 2000. *The Money Pitch: Baseball Free Agency and Salary Arbitration.* Philadelphia, PA: Temple University Press.

Bodvarsson, O. B. and B. King. 1998. "The Value of Arbitration Rights in Major League Baseball: Implication for Salaries and Discrimination." *Quarterly Journal of Business and Economics* 37(1): 65-80.

Domazlicky, B. and P. M. Kerr. 1990. "Baseball Attendance and the Designated Hitter." *American Economist* 34(1): 62-68.

Ferguson, D. G., J. C. H. Jones and K. G. Stewart. 2000. "Competition within a Cartel: League Conduct and Team Conduct in the Market for Baseball Player Services." *Review of Economics and Statistics* 82(3): 422-430.

Fizel, J. 1996. "Bias in Salary Arbitration: The Case of Major League Baseball." *Applied Economics* 28(2): 255-265.

________, E. Gustafson and L. Hadley, Editors. 1996. *Baseball Economics—Current Research.* Westport, CT: Greenwood Publishing.

Horowitz, I. and C. Zappe. 1998. "Thanks for the Memories: Baseball Veterans' End-of-Career Salaries" *Managerial and Decision Economics* 19(6): 377-382.

Kahane, L. and S. Shmanske. 1997. "Team Roster Turnover and Attendance in Major League Baseball." *Applied Economics* 29(4): 425-431.

Kelley, B. P. 1996. *Baseball's Biggest Blunder: The Bonus Rule of 1953-1957.* Lanham, MD: Scarecrow Press.

Knowles, G., K. Shernoy and M. Haupert. 1992. "The Demand for Major League Baseball: A Test of the Uncertainty of Outcome Hypothesis." *American Economist* 36(2): 72-80.

Krautman, A. C. 1999. "What's Wrong With Scully-Estimates of a Player's Marginal Revenue Product?" *Economic Inquiry* 37(2): 369-81.

Leeds, M. and P. von Allmen. 2002. *The Economics of Sports.* New York, NY: Addison Wesley.

Marburger, D. R. 1994. "Bargaining Power and the Structure of Salaries in Major League Baseball." *Managerial and Decision Economics* 15(5): 433-441.

________. 1996. "Racial Discrimination and Long Term Contracts in Major League Baseball." *Review of Black Political Economy* 25(1): 83-94.

McGarrity, J., H. D. Palmer and M. Poitras. 1999. "Consumer Racial Discrimination: A Reassessment of the Market for Baseball Cards." *Journal of Labor Research* 20(2): 247-258.

[1] Postscript: On August 30, 2002, MLB players and owners reached a new collective bargaining agreement without a strike for the first time since 1970. Among other things, the agreement raised the minimum player salary to $300,000, and saved Minnesota Twins and Montreal Expos, at least through 2006.

Miller, P. A. 2000. "A Theoretical and Empirical Comparison of Free Agent and Arbitration-Eligible Salaries Negotiated in Major League Baseball." *Southern Economic Journal* 67(1): 87-104.

Morgan, J. and R. Lally. 1999. *Long Balls, No Strikes: What Baseball Must Do to Keep the Good Times Rolling.* New York, NY: Random House-Crown Publishing.

Richards, D. G. and R. C. Guell. 1998. "Baseball Success and the Structure of Salaries." *Applied Economic Letters* 5(5): 292-296.

Scully, G. W. 1995. *The Market Structure of Sports.* Chicago, IL: The University of Chicago Press.

Schmidt, M. E. and D. J. Berri. 2001. "Competitive Balance and Attendance—The Case of Major League Baseball." *Journal of Sports Economics* 2(2): 145-167.

Sommers, P. M., Editor. 1992. *Diamonds Are Forever: The Business of Baseball.* Washington, D.C.: Brookings Institution Press.

Whitney, J.D. 1988. "Winning Games versus Winning Championships: The Economics of Fan Interest and Team Performance." *Economic Inquiry*: 703-724.

Yilmaz, M. R., S. Chatterjee and M. Habibullah. 2001. "Improvement by Spreading The Wealth: The Case of Home Runs in Major League Baseball." *Journal of Sports Economics* 2(1): 181-193.