6-1-1988

An Actuarial Analysis of the Production Function of Major League Baseball

Michael D. Akers
Marquette University, michael.akers@marquette.edu

Thomas E. Buttross
University of Mississippi

Published version. *Journal of Sport Behavior*. Vol. 11, No. 2 (June 1, 1988): 99-112. Publisher link.

An Actuarial Analysis of the Production Function of Major League Baseball

Michael D. Akers and Thomas E. Buttross
Marquette University and University of Mississippi

The majority of major league baseball production function studies have utilized data (nonstandardized and standardized) from only one baseball season. This research study utilized data from the 1980-1984 baseball seasons. The results indicate that factors contributing to victories fluctuate. Consistent with prior research, this study found batting average, manager's career win-loss record and the strike-out-to-walk ratio to be the most significant factors contributing to victories. Inconsistent with prior research, this study found that fielding average can also be significant.

La majorité des études du rôle de production de "grandes matches de baseball" ont utilisé des données (nonuniformisées et uniformisées) d'une seul saison de baseball. Cette recherche a utilisé les données des saisons de baseball des années 1980-1984. Les resultats montrent que les éléments qui contribuent aux victoires varient. Conformément à la recherche précédente, cette étude a trouvé que le rapport de la moyen des frappeurs, du registre des gains/pertes pendant la carrière du gérant, et du "retirer sur prises pour faire marcher" sont des éléments les plus significants qui contribuent aux victoires. Bien qu'elle est en désaccord avec la recherche précédente, cette étude a trouvé que la moyen d'arrêter les balles est peut être aussi significante.

Management of professional team sports, like management of manufacturing operations, are interested in the factors that influence production. "Economists have been in the forefront of academic efforts in recent years to inform the management of professional team sports with quantitative estimates of the impact of the different factors on performance and attendance at a variety of sports" (Schofield, 1983, p. 196). Of all the sports, major league baseball has received the most attention. The purpose of this paper is to conduct an actuarial analysis of the production function as it relates to major league baseball.

Studies concerning major league baseball have focused on two areas: demand studies and production function studies (Schofield, 1983, pp. 196-197). The statistical method that has been used "almost exclusively" in major league baseball studies is the least squares method of multiple regression (Schofield, 1983, p. 197).

Demand studies have primarily used aggregate attendance as the dependent variable. Independent variables can be grouped into four factors: (1) Economic - price, substitutes or complements, income; (2) Demographic - population size, ethnic mix; (3) Attractiveness of the game - closeness of competition in a particular game, home and visiting team records, special attractions; and (4) Residual preferences - weather, stadium quality, time of the week the game is played (Schofield, 1983, p. 200).

The production function demonstrates the relationships between the inputs and the outputs and gives "the producer information concerning factor shares, marginal productivity, and returns to scale" (Zech, 1981, p. 19). Identifying the production function enhances the efficiency of any operation by allowing the producer to choose the most efficient combination of factors (inputs). Examination of the production function also allows the producer to examine any tradeoffs that exist among the factors.

Major league baseball teams, like other businesses, have a production function. They use inputs (skills of individual players) to produce a product (victories). A production function can be used by major league baseball teams in the following ways: (1) Basis for team member salaries, (2) Determination of which team members should be retained, and (3) Acquisition of talent through trades and free agent market (e.g., teams will need to consider the tradeoffs concerning the question of whether to trade a power hitter for a singles hitter) (Zech, 1981, p. 19).

Production function studies use percentage wins as the dependent variable (output). Independent variables (inputs) that have been used in production function studies include: home runs, batting average, slugging average, stolen bases, fielding average, strike-out-to-walk ratio, earned run average, manager's years as a manager, manager's career win-loss record, a league (American/National) dummy variable, and a contender/noncontender dummy variable (based on games behind the division leader) to measure the influence of management, coaching and morale. Most of the baseball studies that have been conducted have not considered the different units of measure of the independent variables. As a result, Schofield (1983, p. 199) notes that it is not entirely possible to determine which variables are more important due to the nonstandardization of the data.

Scully (1974) found that team slugging average, strike-out-to-walk ratio, and a contender/noncontender dummy variable were significant factors that contributed to victories. The dummy variable used to measure management, coaching and morale should, however, be viewed with skepticism. Contenders were defined as the division winner or teams finishing five games or less behind the division winner. All

other teams in the division were defined as noncontenders. The variable is too general to make any statements concerning managerial efficiency.

One study (Porter & Scully, 1982) examined only the impact of managerial skill on the production process. Porter and Scully (1982) estimated managerial efficiency by manager and by team, managerial marginal revenue product, the rate of change in managerial efficiency over years of experience and relative factor price efficiency using data from the 1961-1980 baseball seasons. A frontier production function and a learning curve were used to estimate managerial efficiency and the rate of change in managerial efficiency. Their results indicated that managerial skill in baseball contributed substantially to the production process.

As noted previously, the problem with these studies has been that the data have not been standardized. One study (Zech, 1981) standardized the variables by using a base team as an index. Zech (1981) found the following variables (listed in order of importance) to be significant factors that contributed to team victories: (1) Batting average, (2) Homeruns, (3) Stolen bases, and (4) Strike-out-to-walk ratio.

Objectives

The objectives of the current study are threefold. The first objective is to conduct an actuarial analysis of the production function of major league baseball using standardized data from the 1980-1984 baseball seasons. The design of this study and the variables utilized are similar to Zech's (1981) study. The study will identify the factors (inputs) that contributed to victories (outputs) for these baseball seasons. Secondly, the results for each of these years will be evaluated to determine if the factors (inputs) that contribute to victories (outputs) are consistent or fluctuating. Thirdly, the results of this study will be compared to prior research.

Method

The Cobb-Douglas production function (McGuigan & Moyer, 1983) with standardized data for the 1980-1984 baseball seasons is utilized to conduct an actuarial analysis of the production function for each major league baseball season. The Cobb-Douglas production function that will be used is of the form: $Q = A X_1^{B1} X_2^{B2} \ldots X_n^{Bn}$ where B_1 through B_n are independent parameters that do not necessarily sum to 1. The Cobb-Douglas production function is an "appealing one for representing the input-output production relationship" (McGuigan & Moyer, 1983, p. 341). While the above equation is nonlinear, it can be transformed into a linear relationship by taking

the natural logarithms of all the variables which results in the following form of the equation: $\ln Q = A + B_1 \ln X_1 + \ln X_2 B_2 + B_2 + \ldots + B_n \ln X_n + e$. When the transformation is made, the parameters (a, B_1, and B_2) can be estimated using least-squares linear regression. This study utilizes the forced entry and stepwise methods of least squares linear regression.

The Cobb-Douglas production function has the benefit of providing information on returns to scale. Specifically, if $B_1 + B_2 + \ldots + B_n > 1$, the organization is experiencing increasing returns to scale. If $B_1 + B_2 + \ldots + B_n < 1$, the organization is experiencing decreasing returns to scale, and if $B_1 + B_2 + \ldots + B_n = 1$, the organization is experiencing constant returns to scale.

Variables

Factors in a baseball production function include the skills of the players that are used in producing victories. There are four main categories of player factors in baseball: (1) Hitting, (2) Running, (3) Defense, and (4) Pitching. Each of these areas contains variables that can be measured to determine their contribution. The following is a discussion of the variables included in the study and the reasons for their inclusion.

Hitting is composed of two aspects: frequency represented by batting average and power represented by homeruns. The type of swing necessary for power is different from the swing necessary for frequency. Thus, a tradeoff exists. As a result, both measures are included in the model.

The best measure of running is stolen bases and thus it is included in the model. The best measure of defense is the fielding average which is also included in the model.

In the area of pitching, two variables are obvious: earned run average (ERA) and strike-out-to-walk ratio. Only the strike-out-to-walk ratio is included in the model. This is the only variable that specifically measures a pitcher's ability. The pitcher's ERA depends not only on the pitcher's ability, but also on the defensive skill of the team. While errors are not considered in the ERA calculation, neither are outstanding defensive plays. Since outstanding defensive plays can affect the ERA calculation, the ERA can be a misleading measure of the pitcher's ability.

The last aspect of the model deals with the contribution of the manager to team victories. This variable is difficult to measure and prior research (Porter & Scully, 1982; Scully, 1974; Zech, 1981) has provided contradictory results. Two measures of manager efficiency (contribution) are used. The first is the manager's career win-loss record and the second is the number of years that the manager has managed

in the major leagues. The career win-loss record should be an indicator of the manager's efficiency. It is also hypothesized that the number of years spent managing would be an indicator of managerial efficiency since poor managers are not retained. The model is tested twice each year. Each time a different measure of the manager's efficiency will be utilized.

Since the model is tested on data from both leagues and since there are differences between the two leagues (designated hitter rule, methods of umpiring, etc.), a dummy variable (National League = 1, American League = 0) is used to control for these differences. The model was tested using data from *The Baseball Encyclopedia* (Reichler, 1985) for the 1980, 1981, 1982, 1983, and 1984 baseball seasons. Testing (for each year separately) was done on a cross-sectional basis since data were used from each of the 26 major league baseball teams. While the size of the samples is small, such samples represent the entire population. Since the variables in the model are all measured in a different unit of measure, the data were indexed to a base team (New York Mets). After the variables were standardized, they were transformed into logarithm form so that least squares regression could be used. Stepwise regression was used in addition to forced entry since there was evidence of multicollinearity.

Actuarial Results

Tables 1 through 5 show the results for each of the seasons (1980-1984) examined in this study. A review of these results provides additional insight concerning the factors that contribute to victories.

In all four models for the 1980 baseball season (Table 1), the F values reveal that all are significant at alpha equals 0.05. No variable was significant in all four models although the strike-out-to-walk ratio was significant in three of the models and the manager's win-loss record was significant in both of the models in which it appeared. There is also evidence that homeruns and batting average are significant factors.

In 1981 (Table 2), all four models were significant at alpha equals 0.05. It is interesting to note that fielding average was significant in all four models. Fielding average had not been a significant factor in prior research. The magnitude of the importance of the fielding average is also surprising. The other surprising result is that there is evidence that the manager's years managed is a significant factor. There is also evidence that homeruns, batting average and the strike-out-to-walk ratio are significant factors.

The F values of the 1982 models (Table 3) reveal that all four models are

Table 1

1980 Baseball Season Models

Variable	Equation 1 (forced entry)	Equation 2 (forced entry)	Equation 3 (stepwise)	Equation 4 (stepwise)
Constant	0.955 (-0.765)	1.059 (0.737)	1.036 (1.112)	1.095 (1.869)
League Dummy	0.049 (1.181)	0.006 (0.099)		
Batting Average	0.133 (0.307)	0.761 (1.299)		1.121* (2.255)
Homeruns	0.112 (1.834)	0.226** (2.902)		0.196* (2.673)
Stolen Bases	0.024 (0.537)	0.067 (1.083)		
Strike-out-to-walk ratio	0.195 (1.478)	0.467* (2.454)	0.255* (2.483)	0.514** (3.592)
Fielding Average	3.490 (0.613)	11.463 (1.424)		
Manager's Career Win-Loss Record	0.697** (4.279)		0.824** (7.969)	
Manager's Years Managed		-0.010 (-0.383)		
Adjusted R-square	0.776	0.552	0.772	0.526
F	13.380**	5.401**	43.281**	10.234**

( ) = t values
** significant at alpha = 0.01
* significant at alpha = 0.05

Table 2

1981 Baseball Season Models

Variable	Equation 1 (forced entry)	Equation 2 (forced entry)	Equation 3 (stepwise)	Equation 4 (stepwise)
Constant	0.841 (-2.075)	0.867 (-1.832)	0.877 (-1.744)	0.924 (-1.112)
League Dummy	0.086 (1.815)	0.087 (2.029)		
Batting Average	1.118* (2.361)	0.907 (2.072)		
Homeruns	0.190* (2.839)	0.161* (2.540)		
Stolen Bases	0.002 (0.043)	- 0.006 (-0.116)		
Strike-out-to-walk ratio	0.212 (1.650)	0.191 (1.612)	0.379* (2.751)	0.319* (2.527)
Fielding Average	25.549** (3.623)	22.844** (3.632)	29.738** (4.644)	24.711** (4.053)
Manager's Career Win-Loss Record	-0.017 (-0.325)			
Manager's Years Managed		0.047 (1.871)		0.070* (2.527)
Adjusted R-square	0.696	0.744	0.572	0.653
F	9.164**	11.365**	17.730**	16.716**

( ) = t values
** significant at alpha = 0.01
* significant at alpha = 0.05

significant at alpha equals 0.05. The only variable that is significant in all four models is batting average. As in 1980, the manager's win-loss record is significant in both of the models in which it appears. The manager's years managed is significant in one of the years. The fact that the strike-out-to-walk ratio was not significant in any model is interesting.

The 1983 models (Table 4), which are all significant at alpha equals 0.05, provide results that are similar to prior research. Both batting average and the strike-out-to-walk ratio are significant in all four models. For the third time, the manager's win-loss record is significant in both of the models in which it appears. There is also evidence that homeruns and stolen bases are significant factors.

The 1984 models (Table 5) are all significant at alpha equals 0.05 and provide some interesting results. While it is not surprising that batting average is significant in all four models, it is surprising that fielding average is significant in all four models. The magnitude of the importance of fielding average is interesting and is similar to the 1981 results. There is also evidence that the manager's win-loss record and homeruns are significant factors. It is particularly noteworthy that the strike-out-to-walk ratio was not significant in any of the models.

A review of the models for all five seasons indicates that no variable is consistently significant. Yet, the most significant variables are batting average and manager's career win-loss record followed by (in order of significance) the strike-out-to-walk ratio, fielding average, homeruns, manager's years managed and stolen bases. Table 6 provides a summary of the number of times that a variable was significant at alpha equals 0.05. Prior research (Scully, 1974; Zech, 1981) found batting average and strike-out-to-walk ratio to be the most important factors contributing to team victories. There is also evidence from prior research (Porter & Scully, 1982; Scully, 1974) that the manager's role is a significant factor contributing to victories.

Since the data have been standardized, the absolute size of the coefficients of the variables indicates which variables contribute the most to victories. The relative importance of the three most significant variables of this study (batting average, manager's career win-loss record, and strike-out-to-walk ratio) can be determined through evaluation of the coefficients. Evaluation of the coefficients indicates that hitters contribute the most to victories, followed by managers and pitchers. If the production function is used to determine salaries, then perhaps hitters should be paid more than managers, and managers should be paid more than pitchers. Prior research (Zech, 1981) also found that hitters contribute more to victories than pitchers. This, however, contradicts the conventional wisdom of baseball in which

Table 3

1982 Baseball Season Models

Variable	Equation 1 (forced entry)	Equation 2 (forced entry)	Equation 3 (stepwise)	Equation 4 (stepwise)
Constant	1.028 (0.449)	0.983 (-0.267)	1.132** (4.024)	1.095* (2.793)
League Dummy	0.038 (0.742)	0.033 (0.636)		
Batting Average	1.616** (3.194)	1.937** (3.947)	1.822** (4.143)	2.207** (4.621)
Homeruns	0.122 (1.511)	0.161 (1.996)		
Stolen Bases	0.059 (1.070)	0.105 (1.958)		
Strike-out-to-walk ratio	0.151 (1.229)	0.199 (1.541)		
Fielding Average	5.727 (0.727)	4.117 (0.512)		
Manager's Career Win-Loss Record	0.439* (2.426)		0.495** (2.865)	
Manager's Years Managed		0.053* (2.242)		
Adjusted R-square	0.605	0.590	0.576	0.449
F	6.472**	6.146**	17.990**	21.354**

( ) = t values
** significant at alpha = 0.01
* significant at alpha = 0.05

Table 4

1983 Baseball Season Models

Variable	Equation 1 (forced entry)	Equation 2 (forced entry)	Equation 3 (stepwise)	Equation 4 (stepwise)
Constant	0.948 (-1.188)	0.911 (-1.695)	0.949 (-1.215)	0.928 (-1.606)
League Dummy	0.035 (1.015)	0.050 (1.211)		
Batting Average	1.056* (2.490)	1.682** (3.797)	1.005* (2.636)	1.420** (3.750)
Homeruns	0.245* (2.802)	0.234* (2.391)		
Stolen Bases	0.077* (2.215)	0.076 (1.810)		
Strike-out-to-walk ratio	0.223* (2.224)	0.331** (3.111)	0.328** (3.487)	0.416** (4.325)
Fielding Average	3.168 (0.558)	-1.642 (-0.265)		
Manager's Career Win-Loss Record	0.416* (2.212)		0.470* (2.474)	
Manager's Years Managed		-0.013 (-0.543))		
Adjusted R-square	0.699	0.624	0.621	0.537
F	9.310**	6.924**	14.683**	15.511**

( ) = t values
** significant at alpha = 0.01
* significant at alpha = 0.05

Table 5

1984 Baseball Season Models

Variable	Equation 1 (forced entry)	Equation 2 (forced entry)	Equation 3 (stepwise)	Equation 4 (stepwise)
Constant	0.862* (-2.782)	0.820** (-3.631)	0.950* (-2.134)	0.899** (-5.326)
League Dummy	0.105 (1.858)	0.114 (1.908)		
Batting Average	1.262* (2.553)	1.379* (2.652)	0.951* (2.216)	1.119* (2.249)
Homeruns	0.208* (2.368)	0.240* (2.668)		
Stolen Bases	0.005 (0.112)	0.032 (0.723)		
Strike-out-to-walk ratio	-0.051 (-0.351)	-0.001 (-0.007)		
Fielding Average	26.743** (3.508)	28.155** (3.480)	20.457* (2.787)	23.041* (2.702)
Manager's Career Win-Loss Record	0.374 (1.658)		0.604** (3.072)	
Manager's Years Managed		0.014 (0.710)		
Adjusted R-square	0.511	0.452	0.456	0.257
F	4.732**	3.941**	7.990**	5.317*

( ) = t values
** significant at alpha = 0.01
* significant at alpha = 0.05

Table 6

Number of Times a Variable Appears Significant at Alpha = 0.05

Variable	1980	1981	1982	1983	1984	Total	Percentage Significant*
Constant	0	0	2	0	4	6	.30
League	0	0	0	0	0	0	.00
Batting Average	1	1	4	4	4	14	.70
Homeruns	1	2	0	2	2	7	.35
Stolen bases	0	0	0	1	0	1	.05
Strike-out-to-walk Ratio	3	2	0	4	0	9	.45
Fielding Average	0	4	0	0	4	8	.40
Manager's Career Win-Loss Record	2	0	2	2	1	7	.70**
Manager's Years Managed	0	1	1	0	0	2	.20**

* Percentage significant based on twenty equations in Tables 1 through 5.
** These two variables only appeared in ten equations in Tables 1 through 5.

pitching is considered 75% of the game.

McGuigan and Moyer (1983, p. 283) define returns to scale as "the proportionate increase in the output of the production process that results from the given proportionate increase in all the inputs." If output (victories) increases more than the proportionate increase in the inputs (independent variables), the production function exhibits increasing returns to scale. If output increases less than the proportionate increase in the inputs, the production function exhibits decreasing returns to scale. Summing the coefficients of the significant variables will indicate if there are increasing or decreasing returns to scale. All models, except equations one and two in 1980, exhibit increasing returns to scale. Equations one and two in 1980 exhibit decreasing returns to scale.

Conclusion

The objectives of this study were to conduct an actuarial analysis of the production function of major league baseball in order to determine the factors that contributed to victories and whether these factors were stable and consistent with

prior research. A limitation of the study is the use of small samples and the results should be interpreted in this context. This study found the following factors to be significant contributors to victories during one or more of the five baseball seasons examined: batting average, homeruns, stolen bases, strike-out-to-walk ratio, fielding average, manager's career win-loss record and manager's years managed. Of these variables, batting average, manager's career win-loss record and strike-out-to-walk ratio were the most consistent with batting average making the largest contribution. The importance of batting average, the strike-out-to-walk ratio and the manager's contribution is consistent with prior research (Porter & Scully, 1982; Scully, 1974; Zech, 1981).

The most significant and interesting result of the study is that fielding average was a significant factor in two of the five seasons. Prior research had not found fielding average to be a significant factor. This study also found, consistent with Zech (1981), that the production function of major league baseball exhibits increasing returns to scale.

The results of this study, while supporting prior research, indicate that factors contributing to victories may fluctuate each season. Since there are several factors that contribute to victories, it is not surprising that the relative importance of these variables could fluctuate each season. Since baseball is such a dynamic game, it is possible that the variables contributing to victories will always fluctuate. Such results could also indicate that there are other variables (e.g., team harmony, relations with top management, player satisfaction) which contribute to victories that cannot be measured statistically. Future research should attempt to operationalize variables such as team harmony and player satisfaction in order to gain a better understanding of the factors that contribute to victories.

References

McGuigan, J. R., & Moyer, R. C. (1983). *Managerial economics.* New York: West.

Porter, P. K., & Scully, G. W. (1982). Measuring managerial efficiency: The case of baseball. *Southern Economic Journal, 48,* 3, 642-650.

Reichler, J. L. (Ed.). (1985). *The baseball encyclopedia* (3rd ed.). New York: Macmillan.

Schofield, J. A. (1983). Performance and attendance at professional team sports. *Journal of Sport Behavior, 6,* 4, 196-206.

Scully, G. (1974, December). Pay and performance in major league baseball. *American Economic Review, 64,* 915-930.

SPSSx User's Guide (2nd ed.). (1986). Chicago: SPSS Inc.

Zech, C. E. (1981, Fall). An empirical estimation of a production function: The case of major league baseball. *American Economist,* 19-23.