Arbitration and Salary Gaps in Major League Baseball

Elizabeth Gustafson*
University of Dayton

Lawrence Hadley
University of Dayton

1990 salary equations for major league baseball players are estimated to assess the relative impacts of performance, service, and market structure on earnings. The current labor market structure guarantees salary arbitration for players with three to six years of major league service. Arbitrated salaries approximate free agent salaries.

The salary gaps between players eligible versus those ineligible for salary arbitration are analyzed for all players and for four subgroups of above average players. The portions of the gap explained by market structure are 25.1 percent ($174,079) and 40.5 percent ($242,965) for all hitters and all pitchers, respectively. These results measure monopsonistic exploitation because arbitration-ineligible players are restricted to playing for a single team and have no recourse for salary disputes.

Finally, there is much less variation in the salaries of arbitration-ineligible players, suggesting an administered salary structure for ineligible players based primarily upon years of service.

INTRODUCTION

Prior to 1973 the labor market for major league baseball players was dominated by the reserve clause. This clause restricted a player's labor market activity to the team holding his contract in the previous season. Because major league teams were not permitted to compete for each other's players, they enjoyed the advantage of a pure monopsonist in dealing with their own players. Economic theory predicts that profit-maximizing monopsonists will exploit workers by paying a wage rate below marginal revenue product. This prediction is based upon the standard textbook analysis of monopsonists who face marginal resource cost functions rising above their labor supply functions. Scully (1974) uses a recursive econometric model to estimate the magnitude of this exploitation. He finds the losses of the baseball players to be considerable.

Since 1973 there have been two dramatic changes in this labor market. First, in 1973 players and owners voluntarily agreed to use binding final offer arbitration to settle salary disputes. At the time, this appeared to be a harmless concession from the owners' viewpoint. The Curt Flood lawsuit challenging the reserve clause recently had been decided in the owners' favor and had been upheld by the Supreme Court.[1] The clause seemed solidly

* We would like to thank Scott M. Fuess and several referees for their helpful comments.

[1] In 1969 Curt Flood had played 12 years with the St. Louis Cardinals and was recognized as a quality outfielder. He was traded to the Philadelphia Phillies against his will and countered with a lawsuit challenging the legality of the reserve clause which allowed the Cardinals to assign his contract to the Phillies. The Supreme Court's opinion (*Flood v. Kuhn, 1972*) decided in favor of major league baseball and its exemption from the antitrust laws. Justice Blackmun recognized baseball's exemption to be an aberration, but

0747-5535/95/1400-0032 $02.50

entrenched as a permanent part of baseball's labor market. The owners had every reason to expect that arbitrators would observe only salaries suppressed by the existing labor market and that these salaries would set the boundaries of their awards.

Second, grievance arbitration overturned the reserve clause in 1975, and major league baseball entered the free agent era. Players currently become eligible for arbitration just before reaching three years of major league service[2] and become eligible for free agency after six years of service.[3]

The combination of free agency and binding salary arbitration has proven to be advantageous for veteran players, as indicated by the upward explosion in their salaries over the past 15 years. Our research has documented a large one time salary increment at the time that players become eligible for arbitration. This salary jump suggests that arbitrators have succeeded in making salary awards that are at least commensurate with those of free agents (Hadley and Gustafson, 1991). The result of this market structure is a salary gap between the players who are eligible for arbitration and those who are ineligible, suggesting that ineligible players may continue to suffer from monopsonistic exploitation.

While the salaries of players eligible for arbitration are definitely higher than the salaries of ineligible players, it could be argued that the difference is due at least partly to improved performance of experienced players and seniority effects rather than exploitation. To isolate the portion of the salary difference that is due to a different reward structure rather than due to different performance, this paper presents a gap analysis for hitters and for pitchers. The portion of the gap between earnings of average eligible and ineligible players explained by their performance and their years of service in the major leagues is justifiable in the context of a competitive labor market. In contrast, the remaining unexplained portion of the gap is interpreted as a measure of exploitation of the arbitration-ineligible players. It should be recognized that important variables omitted from the earnings equations also will contribute to this portion of the gap.

Scoville (1974) hypothesizes that star players will be less exploited than average players because the supply of star players is inelastic. Hill (1985) finds evidence that stars are more exploited than are average players. To pursue this open question we have done a gap analysis for certain groups of above average players.

One may expect an increase in earnings when players become eligible for free agency after six years of service rather than when they become eligible for arbitration at about three years of service. The decision to analyze the arbitration gap rather than the free agency gap is based upon results in our earlier article, which identifies a large one time increase in players' salaries at the time they become eligible for arbitration. Our earlier paper (1991), however, shows no additional significant salary increment after the sixth

---

justified it on the basis of preserving the continuity of earlier decisions, specifically *Federal Baseball v. National League, 1922.*

[2] The number of days of service required for eligibility is based upon a formula from the collective bargaining agreement. In 1990 the new agreement identified the minimum years and days of service to be a number such that 17 percent of the players with at least two but less than three years of service would be eligible. In 1992 this translated into a minimum service requirement of two years, 140 days to qualify for arbitration.

[3] For a thorough description of the labor market in baseball, see Jennings (1990, chapters 1-3 and 8-9), Miller (1990, chapters 8-12), and Scully (1989, chapters 8-9).

year when players become eligible for free agency. This is a logical market outcome if there is competitive bidding for free agents and if arbitrators use free agents' salaries as guidelines in making awards.

To build upon our earlier results, the analysis in this paper focuses on the earnings gap between players who are, or who at one time have been, eligible for salary arbitration versus those players who never have been eligible for arbitration. We group the players on the basis of eligibility for arbitration rather than actual participation in the arbitration process because we believe it is unlikely that arbitration-eligible players would agree to salaries less than those they expect to win via the arbitration process.

THE ANALYSIS OF AN EARNINGS GAP FOR BASEBALL PLAYERS

Underlying our analysis is a single equation model of earnings for major league baseball players. Specifically, we estimate the following earnings regression equation separately for hitters and for pitchers:

$$(1)\ \ln \widehat{salary} = \hat{\alpha}_N + \hat{\beta}_N X + \hat{\alpha}_E D + \hat{\beta}_E DX + \hat{\gamma} S$$

where:

- S = A vector of the years of service variables, SERVE and SERVE^2;
- D = A dummy variable equal to one for players who have been eligible for salary arbitration in 1990 or earlier; and
- X = A vector of all other explanatory variables in the model.

The subscripts E and N refer to players who have been eligible and who have never been eligible for salary arbitration, respectively. Vectors of estimated regression coefficients for the various explanatory variables and constant term are symbolized by $\hat{\beta}_N$, $\hat{\alpha}_N$, $\hat{\beta}_E$, $\hat{\gamma}$, and α_N. All of the variables are defined in Table 1.

Following Mincer (1974), the form of the earnings equation is semi-logarithmic with the natural log of annual earnings (lnSALARY) as the dependent variable. This form of the model imposes a constant rate of return in salary to changes in the explanatory variables rather than a constant absolute salary return.

Also following Mincer, the square of years of service is entered as an independent variable to allow for the possibility of diminishing returns to experience. The SERVE and SERVE^2 variables are mainly expected to capture any institutional effects that seniority may have upon salaries. The variables also may capture the physical effects of aging because SERVE is a close proxy for a player's age. Expectations about a player's future performance depend upon the player's position in the typical career cycle of an athlete as well as on current performance. Finally, these variables also may capture the player's mastery of the intangible aspects of the game that are learned with experience but which cannot be measured easily by performance variables.

Table 1—Variables in the Earning Equations

Variable	Definition
Variables Common to Both Salary Equations	
lnSALARY	= The natural logarithm of a player's 1990 salary
SERVE	= A player's years of major league service
$SERVE^2$	= SERVE squared
CONT91, CONT92	= Binary variables equal to 1 for players with contracts of two years and three or more years, respectively
BLACK	= A binary variable equal 1 for black players
HISPAN	= A binary variable equal 1 for hispanic players
NL	= A binary variable equal 1 for National League players
TPCT	= The player's team's winning percentage in 1989
ATTEND	= The player's team's total home attendance in 1989 (measured in hundred thousands)
SMSA	= The population of the metropolitan area of the team (measured in millions)
Variables Unique to the Hitters' Equation	
HITS	= Career hits divided by career at bats (career batting average)
HR	= Career home runs divided by career at bats
RBI	= Career runs batted in divided by career at bats
RUNS	= Career runs scored divided by career at bats
SB	= Career stolen bases divided by career at bats
SOH	= Career strike outs divided by career at bats
BBH	= Career base on balls divided by career at bats
AB	= Average at bats per year of major league service
CATCH, FIRST, SECOND, THIRD, SS, and DH	= Six binary variables equal 1 for players whose primary fielding positions are catcher, first baseman, second baseman, third baseman, shortstop, and designated hitter, respectively
Variables Unique to the Pitchers' Equation	
WINS	= Career wins divided by career innings pitched
LOSSES	= Career losses divided by career innings pitched
SOP	= Career strike outs divided by career innings pitched
BBP	= Career base on balls divided by career innings pitched
SAVE	= Career saves divided by career innings pitched
CGAME	= Career completed games divided by career innings pitched
ERA	= Career earned run average
IP	= Average innings pitched per year of major league service
START	= A binary variable equal 1 for starting pitchers

Salary data are from *The Sporting News*, April 23, 1990, pp. 22-23. Baseball data are from *The Baseball Encyclopedia*, 1990, *The 1991 American League Red Book*, and *The 1991 National League Green Book*

The vector of explanatory variables, X, includes career performance variables specific to each player, variables measuring player attributes, and team-specific variables. There are eight player performance variables in the hitters' equation, and they are defined in terms of a player's career totals relative to his career at bats. They include HITS (hits per at bats or career batting average), HR (home runs per at bat), RBI (runs batted in per at bat), RUNS (runs scored per at bat), SB (stolen bases per at bat), SOH (strike outs per at bat), BBH (bases on balls per at bat), and AB (average at bats per year of major league service). Except for SOH, larger values of all of these variables are associated with better performance. Therefore, the regression coefficients in the salary equation are expected to be positive, except for the coefficient of SOH which is expected to be negative.[4]

There are also eight player performance variables in the pitchers' equation, and they are defined in terms of a player's career totals relative to his career innings pitched. They include WINS (wins per innings pitched), LOSSES (losses per innings pitched), SOP (strikeouts per innings pitched), BBP (bases on balls per innings pitched), SAVE (saves per innings pitched), CGAME (complete games per innings pitched), ERA (career earned run average), and IP (average innings pitched per year of major league service). Regression coefficients for all of these variables except LOSSES, ERA, and BBP are expected to be positive because higher values indicate superior performance. The coefficients of LOSSES, ERA, and BBP are expected to be negative.[5]

All performance variables for hitters and pitchers are measured over the player's career rather than the previous season. On the surface, a marginal measure such as the previous season's performance may appear to be preferred. Players are compensated, however, for their unobserved performance in the coming season. It seems more logical to use the career history of the player to predict future performance than to use only information from the previous season.

Variables in the equation measuring player attributes include two binary variables related to the length of the players' contracts (CONT91 equals one for all players on two year contracts and CONT92 equals one for all players on contracts of three or more years), binary race variables BLACK and HISPAN (HISPAN equals one for all Latin-born and US-born Hispanic players),[6] binary variables related to the players' fielding positions in the hitters' equation, and a binary variable in the pitchers' equation identifying starting pitchers. CONT91 and CONT92 are expected to have positive coefficients because superior players are more likely to be awarded long-term contracts, and the coefficients of the race variables are expected to be negative if racial discrimination manifests itself in lower salaries for minority players. The coefficients for fielding positions that require exceptional skill with the glove are expected to be positive. These may include the variables identifying catchers, shortstops, and perhaps second and third basemen. The coefficient for

[4] A wide range of performance variables is introduced into the hitters' salary equation in order to capture various aspects of performance on the field. Hitters exhibit varying degrees of skill in three basic areas:

- The ability to get on base;
- The ability to hit with power; and
- The ability to run.

A player with exceptional ability to get on base would have high values for HITS, RUNS, and BBH and a low value for SOH. In the ability to hit with power, an exceptional player would have high values for HR, RBI, and SOH (the downside of power hitting). In the ability to run, an exceptional player would have high values for RUNS and SB.

[5] A wide range of performance variables is introduced into the pitchers' salary equation to capture various aspects of performance. Basically, there are two types of pitchers. One uses velocity to retire hitters, and the other combines location control with change of pitching speed to retire hitters. The first type would have high values for SOP while the second would have low values for BBP. WINS and SAVE are included in our equation because they are the primary measure of success for starting pitchers and relief pitchers, respectively. CGAME is a secondary measure of success for starting pitchers, and LOSSES capture the downside of all pitching. Finally, all exceptional pitchers would be expected to have low values for ERA which is the number of runs charged against the pitcher per nine innings pitched.

[6] The racial or ethnic identity of most players is well-known. For players with ambiguous identities, phone calls were made to the front offices of their teams. All front offices were cooperative.

starting pitchers also is expected to be positive because a superior pitcher is usually the team's starter.

The team-specific variables in the equations include a binary variable identifying National versus American League teams and variables measuring the teams' winning percentages in the 1989 season, the teams' home attendance in 1989, and the size of the metropolitan area in which the teams play their home games. The league variable is included because only the American League allows a designated hitter to bat in place of the pitcher. There is a potential, but uncertain, impact on salaries that is controlled by including this variable. High values for the other three team variables are associated positively with a team's ability to pay. More successful teams may pay their players above market salaries in an effort to attract the most skilled players and increase the probability of winning championships. Therefore, the coefficients of these three variables are expected to be positive.

Equation (1) expresses different earnings equations for the arbitration-eligible and ineligible groups of players. The years of service variables have the same coefficients for both groups, but all other coefficients vary by group. Equation (2) shows the predicted lnSALARY for the average player who is ineligible for arbitration as:

$$(2)\ \overline{\widehat{\ln SALARY_N}} = \hat{\alpha}_N + \hat{\beta}_N\bar{X}_N + \hat{\gamma}\bar{S}_N\ .$$

For the average eligible player it is:

$$(3)\ \overline{\widehat{\ln SALARY_E}} = (\hat{\alpha}_N + \hat{\alpha}_E) + (\hat{\beta}_N + \hat{\beta}_E)\bar{X}_E + \hat{\gamma}\bar{S}_E$$

where $\bar{X}_N$, $\bar{S}_N$, $\bar{X}_E$, and $\bar{S}_E$ are variable averages for the two groups.

We define the earnings gap as:

$$(4)\ G \equiv \overline{\widehat{\ln SALARY_E}} - \overline{\widehat{\ln SALARY_N}}\ .$$

We can calculate components of the gap due to the differences in average X and S as well as to differences in the coefficients of X. With some manipulation, G can be written as:

$$(5)\ G = [(\hat{\beta}_N + \hat{\beta}_E)(\bar{X}_E - \bar{X}_N)] + [\hat{\alpha}_E + \hat{\beta}_E\bar{X}_N] + \hat{\gamma}[\bar{S}_E - \bar{S}_N].$$

Defining $\Delta\bar{X} \equiv (\bar{X}_E - \bar{X}_N)$ and $\Delta\bar{S} \equiv (\bar{S}_E - \bar{S}_N)$, we can show the components of the gap:

$$(6)\ 1 = [(\hat{\beta}_N + \hat{\beta}_E)\,\Delta\bar{X}]/G + (\hat{\alpha}_E + \hat{\beta}_E\bar{X}_N)/G + \hat{\gamma}\,\Delta\bar{S}/G.$$

where $\hat{\alpha}_E$ and $\hat{\beta}_E$ are the differences in coefficients between the E and N groups.

The first component, $[(\hat{\beta}_N + \hat{\beta}_E)\,\Delta\bar{X}]/G$, is the fraction of the gap due to differences in average attributes evaluated using the group E coefficients. The second component, $(\hat{\alpha}_E$

$+ \hat{\beta}_E\bar{X}_N)/G$, is the fraction of the gap due to differences in the coefficients applied to the ineligible group's attributes. The third component, $\hat{\gamma}\,\Delta\bar{S}/G$, is the fraction due to different average years of service.

If the second component is close to zero, then different earnings for eligible and ineligible players could be explained in terms of seniority and differences in the skills of the two groups. The larger is the second component, the more the ineligible players may be exploited by the reserve clause. Specifically, we define the degree of exploitation as $(\hat{\alpha}_E + \hat{\beta}_E\bar{X}_N)/G$—the fraction of the gap accounted for by the differences in the returns to the performance and attribute variables. This definition assumes that arbitrators' awards approximate competitive market salaries.

We also assume that the entire portion of the gap unexplained by differences in mean values is due to monopsonistic exploitation rather than to omitted variables. For example, the ability of ineligible players to sustain their observed level of performance is more uncertain than for eligible players; this may be an omitted variable providing an explanation of lower returns to performance. We believe, however, that this effect is captured at least partially by SERVE and SERVE^2 in the earnings equations.

Scully (1974) and Hill (1985) have analyzed the labor-market exploitation of baseball players using an alternative method. They define exploitation as the difference between a player's marginal revenue product (MRP) and his earnings as a percent of MRP. This method requires an indirect estimation of MRP based upon the assumption that a hitter's percent of the team's total at bats (pitcher's percent of the team's total innings pitched) and his performance on the field determine that hitter's (pitcher's) contribution to team revenue. Our relative measure of exploitation avoids this assumption because its estimation requires only observed salary, performance, and attribute data. It also identifies the amounts (percentages) of exploitation accounted for by specific independent variables in the earnings equations. Our method must assume, however, that the labor market for eligible players is competitive and that, therefore, their earnings are equal to their marginal revenue products. We believe this assumption is justified on the basis of the active bidding for free agent players between 1988 and 1990 and the influence these market-determined salaries have had upon arbitrators' awards (Hadley and Gustafson, 1991, pp. 119-120).

Because the difference in performance and attributes is weighted by the eligible group's coefficients rather than by the ineligible group's coefficients, this measure of exploitation may underestimate actual exploitation. This practice is standard, however, in wage-gap analysis.[7] Oaxaca's (1973) method for earnings gap analysis estimates separate regression equations for two subgroups. In our case, the subgroups are eligible and ineligible baseball players. We modify the conventional method by requiring the coefficients for the years of service variables, SERVE and SERVE^2, to be the same for the eligible and ineligible groups. Thus, the other coefficients are allowed to vary by group by using interaction terms, but there are no interaction terms for SERVE and SERVE^2. Our earlier research (Hadley and Gustafson, 1991) shows that in addition to the jump in earnings that

[7] The conventional method developed by Oaxaca (1973) has been used continually to analyze earnings gaps between various segments in the labor market. The literature on the application of this methodology to the male-female earnings gap is summarized in Gunderson (1989).

occurs at three years of service, the slopes of the earnings-service profiles for hitters and pitchers are much steeper in the early years of service than in the later years. The use of the squared service variable allows for the decrease in slope as service increases; this is part of the typical earnings-service profile for most labor markets. If we estimate separate coefficients for SERVE and SERVE^2 for the eligible and ineligible groups, we clearly would observe large differences in these coefficients that would be misinterpreted as part of the exploitation. Therefore, the coefficients of the years of service variables (and, thus, the earnings-service profile) are estimated for the whole data set rather than separately for the two groups.

EMPIRICAL ESTIMATION OF THE EARNINGS EQUATION

We estimate the earnings equations for hitters and pitchers using ordinary least squares. To investigate possible heteroscedasticity, we graph the squared residuals relative to each explanatory variable. We also run regressions of the squared residuals on individual explanatory variables. For the hitters' equation, the regressions on SERVE, HR, and RBI show significance. For the pitchers' equation, regressions on ERA and SOP are significant. Using these groups of variables to explain the variance of the disturbances, we perform the Breusch-Pagan test (Breusch and Pagan, 1979). Chi-square test statistics of 60.63 and 40.83 for hitters and pitchers respectively indicate significant heteroscedasticity at the .05 level.

To correct for heteroscedasticity, we reestimate the equations using weighted least squares. We use the inverses of the predicted standard deviations of the disturbances as the weights. We regress the logs of the squared residuals on SERVE, HR, and RBI for hitters and on ERA and SOP for pitchers, on the squares of these variables, and on their cross-products. From these equations, we obtain the predicted standard deviations of the disturbances.[8] The weighted least squares results for hitters and pitchers are presented in Tables 2 and 3 respectively.

Observations for the hitters' regression equation include all nonpitching players for whom salary data are available and who had at least 100 career at bats in the major leagues prior to the 1990 season. Observations for the pitchers' regression equation include all pitchers for whom salary data are available and who pitched at least 50 innings in the major leagues prior to the 1990 season.

SALARY is measured with annual salary data for the 1990 baseball season that were obtained from *The Sporting News* (April 23, 1990, pp. 22-23). These data include base salaries and signing bonuses. The performance data were obtained from the eighth edition of *The Baseball Encyclopedia* (1990), *The 1991 American League Red Book* (1991), and *The 1991 National League Green Book* (1991).

EMPIRICAL RESULTS

Tables 2 and 3 show the estimates of equation (1) for hitters and pitchers, respectively. The gap analysis and level of relative exploitation based on these results are

---

[8] The method used to obtain weights to correct for heteroscedasticity is outlined in detail in Ramanathan (1992, pp. 349-351). We took antilogs from the predicted logs of variances of the disturbances to get predictions of the variances.

Table 2—Regression Analysis of the Arbitration Earnings Gap: Hitters (Dependent Variable = lnSALARY)#

	Means		Regression Coefficient		
	Players Not Eligible for Arbitration	Players Eligible for Arbitration	Players Not Eligible for Arbitration	Interaction	Players Eligible for Arbitration
SERVE	1.673	7.348	0.259***		0.259***
	(0.718)	(3.577)	(0.0291)		(0.0291)
SERVE^2	3.310	66.737	-0.0134***		-0.0134***
	(2.538)	(67.331)	(0.00185)		(0.00185)
HITS	0.244	0.263	0.468	1.600	2.068
	(0.029)	(0.023)	(1.034)	(2.192)	(1.933)
HR	0.017	0.025	8.209***	-11.731***	-3.522
	(0.013)	(0.016)	(3.057)	(5.718)	(4.829)
RBI	0.100	0.119	4.792***	2.433	7.225***
	(0.029)	(0.031)	(1.534)	(2.658)	(2.169)
RUNS	0.121	0.131	-1.621	8.761***	7.140***
	(0.028)	(0.024)	(1.484)	(3.120)	(2.743)
SB	0.024	0.027	2.287	-2.426	-0.139
	(0.025)	(0.028)	(1.581)	(2.368)	(1.766)
SOH	0.178	0.159	-0.0125	-0.0297	-0.0422
	(0.063)	(0.055)	(0.465)	(0.771)	(0.616)
BBH	0.079	0.094	0.803	-1.165	-0.362
	(0.031)	(0.033)	(0.791)	(1.264)	(0.985)
AB	292.916	389.511	0.000787***	0.00214***	0.0029***
	(132.050)	(129.187)	(0.000238)	(0.000354)	(0.000264)
CATCH	0.168	0.147	-0.00781	0.512***	0.504***
	(0.375)	(0.355)	(0.0689)	(0.109)	(0.0844)
FIRST	0.056	0.095	-0.0173	0.254*	0.237***
	(0.231)	(0.294)	(0.106)	(0.137)	(0.0866)
SECOND	0.088	0.091	0.200*	-0.235	-0.0350
	(0.284)	(0.289)	(0.109)	(0.155)	(0.111)
THIRD	0.080	0.083	-0.0571	0.166	0.109
	(0.272)	(0.277)	(0.0800)	(0.116)	(0.0838)
SS	0.104	0.099	0.206*	-0.0342	0.172*
	(0.306)	(0.300)	(0.109)	(0.146)	(0.0971)
DH	0.008	0.036	0.0769	0.0310	0.108
	(0.089)	(0.186)	(0.191)	(0.251)	(0.163)
CONT91	0	0.131	0.0	0.0255	0.0255
	(0)	(0.338)	(0.0)	(0.0891)	(0.0891)
CONT92	0.008	0.143	-0.432**	0.532**	0.100
	(0.089)	(0.351)	(0.211)	(0.226)	(0.0865)
BLACK	0.248	0.294	-0.0325	0.133	0.101
	(0.434)	(0.456)	(0.0641)	(0.0911)	(0.0650)
HISPAN	0.200	0.151	0.0861	0.153	0.239***
	(0.402)	(0.359)	(0.0754)	(0.105)	(0.0726)
NL	0.456	0.456	0.00684	0.0653	0.0721
	(0.500)	(0.499)	(0.0506)	(0.0723)	(0.0516)
TPCT	0.499	0.499	-1.637***	1.571*	-0.0662
	(0.056)	(0.061)	(0.648)	(0.851)	(0.555)
ATTEND	20.550	21.267	0.0179***	-0.00165	0.0163***
	(6.046)	(6.091)	(0.00685)	(0.00884)	(0.00560)
SMSA	5.751	5.630	-0.00581	-0.00463	-0.0104*
	(4.826)	(4.738)	(0.00582)	(0.00839)	(0.00602)
CONSTANT			11.070***	-2.439***	8.631***
			(0.435)	(0.622)	(0.456)
SAMPLE N	125	252			

R squared 0.879 (Entire model including interaction terms)
F statistic 5.39*** (Test for difference between regressions for players eligible and not eligible for arbitration.)
* (**) (***) indicates statistical significance at the .10 (.05) (.01) level using two-tailed t-tests
#For definitions of variables, see Table 1. Standard deviations and standard errors in parentheses

Table 3—Regression Analysis of the Arbitration Earnings Gap: Pitchers (Dependent Variable = lnSALARY)#

	Mean		Regression Coefficient		
	Players Not Eligible for Arbitration	Players Eligible for Arbitration	Players Not Eligible for Arbitration	Interaction	Players Eligible for Arbitration
SERVE	1.566	6.813	0.157***		0.157***
	(.711)	(3.477)	(0.0246)		(0.0246)
$SERVE^2$	2.952	58.433	-0.0066***		-0.0066***
	(2.432)	(67.759)	(0.00112)		(0.00112)
WINS	0.049	0.058	4.030*	-0.0490	3.981
	(0.018)	(0.010)	(2.167)	(3.565)	(2.815)
LOSSES	0.060	0.056	-2.080	3.726	1.646
	(0.021)	(0.010)	(1.850)	(3.603)	(3.095)
SOP	0.645	0.657	0.160	0.879***	1.039***
	(0.180)	(0.152)	(0.194)	(0.277)	(0.201)
BBP	0.413	0.385	-0.354	0.534*	0.180*
	(0.114)	(0.079)	(0.305)	(0.320)	(0.095)
SAVE	0.022	0.043	2.191***	-0.677	1.514***
	(0.059)	(0.075)	(0.490)	(0.736)	(0.553)
CGAME	0.008	0.015	0.852	-4.402	-3.550
	(0.009)	(0.014)	(4.781)	(6.447)	(4.352)
ERA	4.064	3.688	-0.0155	-0.296***	-0.312***
	(0.904)	(0.525)	(0.0528)	(0.0949)	(0.0789)
IP	118.541	139.182	0.00361***	0.00507***	0.0087***
	(46.561)	(50.363)	(0.00123)	(0.00175)	(0.00123)
START	0.460	0.516	-0.0885	0.0298	-0.0587
	(0.501)	(0.501)	(0.0967)	(0.132)	(0.0903)
CONT91	0	0.120	0.0	0.277***	0.277***
	(0)	(0.325)	(0.0)	(0.0896)	(0.0896)
CONT92	0	0.120	0.0	0.242***	0.242***
	(0)	(0.325)	(0.0)	(0.0913)	(0.0913)
BLACK	0.053	0.038	-0.0255	-0.320*	-0.345***
	(0.225)	(0.192)	(0.127)	(0.172)	(0.116)
HISPAN	0.071	0.092	-0.0875	0.0150	-0.0725
	(0.258)	(0.290)	(0.131)	(0.159)	(0.0900)
NL	0.407	0.473	-0.0328	-0.0128	-0.0456
	(0.493)	(0.501)	(0.0657)	(0.0847)	(0.0534)
TPCT	0.492	0.511	-0.844	1.751	0.907
	(0.059)	(0.054)	(0.850)	(1.071)	(0.648)
ATTEND	20.663	22.240	0.0142*	-0.0268***	-0.0126**
	(6.682)	(5.701)	(0.00752)	(0.00977)	(0.00625)
SMSA	5.473	5.926	-0.00347	0.0129	0.00943*
	(4.432)	(5.183)	(0.00733)	(0.00915)	(0.00547)
CONSTANT			11.489***	-0.0920	11.397***
			(0.466)	(0.681)	(0.516)
SAMPLE N	113	184			

R squared 0.858 (Entire model including interaction terms)
F statistic 6.48*** (Test for difference between regressions for players eligible and not eligible for arbitration.)
* (**) (***) indicates statistical significance at the .10 (.05) (.01) level using two-tailed t-tests
#For definitions of variables, see Table 1. Standard deviations and standard errors in parentheses

presented in Table 4. To test Scoville's (1974) hypothesis, gap analysis also is performed for subgroups of above average performance players to determine whether stars are exploited to a different degree than average players. These results are also in Table 4.

In Tables 2 and 3, the first and second columns show the mean values of the variables for the arbitration-ineligible and eligible players, respectively. The differences in

Table 4—Summary of Arbitration Earnings Gaps (Gaps are shown in dollars in parentheses.)

Group	Predicted Salary of Mean Ineligible Player	Predicted Salary of Mean Eligible Player	Percent of Gap Explained by Differences in Performance Variables	Percent of Gap Explained by Other Variables	Percent of Gap Explained by Differences in Regression Coefficients (Exploitation)	Percent of Gap Explained by Differences in Years of Service
All Hitters	$148,494	$848,009	28.56 ($198,539)	1.35 ($9,358)	25.10 ($174,079)	44.99 ($317,539)
Finesse Hitters, HITS > .257	$189,986	$1,013,316	15.47 ($124,702)	-0.82 (-$6,592)	56.14 ($468,070)	29.21 ($237,150)
Power Hitters, HR > .022	$165,270	$1,225,161	37.48 ($397,752)	-1.87 (-$19,387)	23.88 ($249,934)	40.51 ($431,592)
All Pitchers	$157,102	$754,164	21.67 ($128,490)	3.91 ($23,017)	40.49 ($242,965)	33.93 ($202,590)
Finesse Pitchers, ERA < 3.831	$150,864	$917,402	23.40 ($176,606)	2.58 ($19,329)	52.76 ($410,356)	21.26 ($160,247)
Power Pitchers, SOP > .653	$166,189	$897,540	24.26 ($174,689)	4.61 ($32,987)	55.30 ($410,095)	15.83 ($113,580)

these means determine the first component of the gap which is interpreted as the market-justified differences in earnings between the two groups. The third column reports the regression coefficients for the ineligible group of players. The fourth column shows the coefficients of the interaction terms which are added to the ineligible players' coefficients to obtain the regression coefficients for the arbitration-eligible players. These arbitration-eligible coefficients are reported in the fifth column.

As an example of interpreting these tables, consider the RBI variable in Table 2. The average eligible player has a 0.119 career RBI average, and the average ineligible player has only a 0.100 RBI average. The coefficient in the third column indicates that for ineligible players a one unit change in RBI yields a return in lnSALARY of 4.792. This translates to approximately a 479.2 percent change in salary or to a 4.79 percent change for a more reasonable change in RBI of 0.01. Eligible players receive a larger return of 7.225. There is a difference between the two groups in the return to RBI of 2.433. We see that although the eligible players have a better performance in their RBIs, they are also more highly paid for their performance in terms of RBIs.

All significant coefficients for the service and performance variables have the expected signs except BBP in the pitchers' equation (significantly positive at the .10 level only for arbitration-eligible players). Other unexpected signs on significant coefficients appear on ATTEND in the pitchers' equation, SMSA for hitters' eligible for arbitration, and TPCT and CONT92 for hitters ineligible for arbitration.

The significantly negative difference between the constant terms for the eligible and ineligible hitters seems inappropriate at first glance, but it easily is explained. The interaction terms allow changes in the slopes of the earnings functions in addition to changes in the intercepts. The slope is steeper for eligible players in the hitters' equation, thereby forcing the intercept to a lower level as the two lines cross. As a result, a comparison of eligible and ineligible players with the same performance still shows the eligible players with greater earnings as expected.

An F-test shows the overall equation for the eligible players is significantly different from the equation for the ineligible players at the .01 level. (See Tables 2 and 3 for the F statistics.) The equations use a large number of performance variables rather than one or two summary measures to observe subtle differences between the groups' performances. The high correlation among these variables leads to the insignificance of individual coefficients but does not harm the ability of the model to predict salaries.[9]

Table 4 presents the results of the gap analysis for all hitters and pitchers as well as for four subgroups of players exhibiting above average performance standards. The two subgroups of hitters include hitters with a career batting average (HITS) greater than the sample mean of 0.257 (designated as *finesse hitters* in the following discussion) and hitters with a career home run average (HR) greater than the sample mean of 0.022 (designated as *power hitters*).[10] The two subgroups of pitchers include those with a career ERA less than the sample mean of 3.831 (designated as *finesse pitchers*) and pitchers with a career strike out average (SOP) greater than the sample mean of 0.653 (designated as *power pitchers*).[11] (Due to space constraints, the regression results for these four subgroups are not reported in this paper. They are available from the authors upon request.)

In Table 4, it is useful to divide the X variables into those measuring player performance and all others such as race, team performance, and fielding position. The first component of the gap can be divided into a portion due to differences in mean player performance and a portion due to differences in the means of the other variables.

For both hitters and pitchers, the portion of the gap explained by differences in the regression coefficients is substantial (25.10 and 40.49 percent, respectively). If all players sold their services under the same set of market rules, then the returns to player attributes, as measured by these coefficients, should be the same for all players. Therefore, the portion of the gap explained by differences in the regression coefficients is the unjustified portion that is attributable to market restrictions placed upon young players by the reserve clause. The actual dollar amounts of these portions of the two gaps are $174,079 and

[9] We also estimated a more parsimonious form of each equation using only two performance variables (the ratio of strike-outs to base on balls and innings pitched for the pitchers' equation and slugging percent and at bats for the hitters' equation). This specification results in significant coefficients for all performance variables and their interactions. The explanatory power of the equations is somewhat lower, and results for other explanatory variables are similar to those reported in Tables 2 and 3. We choose to conduct the gap analysis with the larger model to allow discussion of differences in various components of players' performances.

[10] We define *finesse hitting* as the abilities to make contact at the plate (as opposed to striking out frequently) and to hit consistently to all sections of the field. These abilities translate into superior performance in getting on base. We use above average values for HITS to identify above average finesse hitters. We define *power hitting* as the ability to drive the baseball long distances. We use above average values of HR to identify hitters with above average power-hitting ability.

[11] We define *finesse pitching* as the ability to control the location of a pitch and to mix pitches with varying velocity. We use below average values of ERA to identify above average finesse pitchers. We define *power pitching* as the ability to throw the baseball with high velocity and thus to overpower and strike out hitters. We use above average values of SOP to identify pitchers with above average power-pitching ability.

$242,965 for hitters and pitchers, respectively. It appears that the reserve clause is somewhat more unfair to ineligible pitchers than to hitters.

The portion of the earnings gap explained by years of service is also substantial (44.99 percent for hitters and 33.93 percent for pitchers). In contrast, the portion of the gap explained by the other variables (race, team performance, position variables, etc.) is small (1.35 percent for hitters and 3.91 percent for pitchers).

The individual performance variables generally are important determinants of the earnings gaps between arbitration-eligible veterans and arbitration-ineligible young players. For all hitters, the overall average performance of the arbitration-eligible veterans is superior, and this explains 28.56 percent of the earnings gap. A smaller percent of the gap is explained by performance differences in the case of finesse hitters (15.47 percent), and a larger percent is explained for power hitters (37.48 percent). Looking at the specific performance variables that explain the earnings gap of all hitters, AB, RBI, RUNS, and HITS account for 16.23, 7.88, 4.10, and 2.62 percent of the gap, respectively. The other performance variables explain negligible portions of this gap. For the finesse hitters' gap, AB and RBI explain 9.18 and 6.64 percent, respectively. For the power hitters' gap, RBI, AB, HITS, and RUNS explain 14.60, 11.73, 7.52, and 5.46 percent, respectively. The other performance variables explain negligible portions of these gaps.

One important result of the gap analysis reported in Table 4 is that the subgroup of ineligible finesse hitters is exploited more by the reserve clause than the overall group of ineligible hitters (56.14 percent versus 25.10 percent). Our analysis indicates that $468,070 of the total earnings gap for hitters with HITS greater than .257 is explained by differences in the regression coefficients. In sharp contrast, the dollar amount of the gap explained by the regression coefficients is $174,079 for all hitters. For arbitration-ineligible power hitters (HR > 0.022), the percent of the gap due to differences in the regression coefficients is approximately the same as for all hitters (23.88 and 25.10). In dollar terms, however, the portion of the power hitter's gap due to exploitation is $249,934 which is $75,855 larger than corresponding value for all hitters. This absolute dollar gap is larger for the power hitters because the mean salary of arbitration-eligible power hitters is large compared to that of all arbitration-eligible hitters.

For all pitchers, the overall performance of eligible veteran players is superior to the ineligible young players, and this explains 21.67 percent of the earnings gap. For finesse pitchers, 23.40 percent of the earnings gap is explained by the performance variables, while 24.26 percent is explained for power pitchers. Looking at the specific performance variables that explain the earnings gap of all pitchers, IP, ERA, WINS, and SAVE account for 11.42, 7.47, 2.28 and 2.03 percent, respectively. For the finesse pitchers' gap, IP, ERA, SAVE, and WINS explain 16.83, 5.60, 2.00, and 1.68 percent respectively. For the power pitchers' gap, IP, ERA, and SAVE explain 16.60, 8.02, and 2.52 percent, respectively.

Both subgroups of ineligible pitchers (those with ERA less than 3.831 and SOP greater than 0.653) are exploited more by the reserve clause than the overall group of ineligible pitchers. The portion of the finesse pitchers' gap explained by the regression coefficients is 52.76 percent compared to 40.49 percent for all pitchers. The portion for power pitchers is even higher at 55.30 percent. For all pitchers, the absolute value of the portion of the gap explained by the regression coefficients is $242,965. The 52.76 percent for finesse pitchers translates into $410,356, and the 55.30 percent for power pitchers translates into $410,095.

CONCLUSIONS

We estimate earnings functions for major league baseball players which indicate that players are rewarded according to performance and years of service, but players with less than three years of service earn substantially less than veteran players. This is the expected outcome due to the monopsonistic position of the teams when dealing with these players. Furthermore, the results reported in Table 4 indicate that three of the four subgroups of above average ineligible players are more exploited by the reserve clause than the average ineligible players.

Hill (1985) also finds that the degree of exploitation increases for players with above average performance statistics. This result is inconsistent with Scoville's (1974) assertion that above average players should be less exploited than average players because their supply is highly inelastic.

It is not surprising that supply elasticity has little to do with the salaries of arbitration-ineligible players. Given the rules of the market, the demand for these players is determined unilaterally by the single baseball club which holds their contract. There are restrictions negotiated in the collective bargaining agreement such as the minimum salary, but our results indicate that the dominant reality is the monopsonistic market structure faced by ineligible players.

It is as if the major league clubs have agreed (implicitly or explicitly in secret) to rough salary scales structured on the basis of years of service for most arbitration-ineligible players. In 1990 nearly all players without major league service were paid $100,000 which was the minimum salary negotiated in the collective bargaining agreement. Salaries for players with one and two years of service paralleled each other closely. Only after three years of service do salaries diverge on the basis of performance.

Evidence supporting the idea of a structured salary scare for arbitration-ineligible players can be found in the coefficients of variation (standard deviation/mean) of players' salaries by service cohorts. These coefficients for cohorts of players with up to 12 years

Table 5—Coefficients of Variation of Salary by Service Cohorts

	Hitters			Pitchers		
SERVE	Mean Salary	Coefficient of Variation	Sample	Mean Salary	Coefficient of Variation	Sample
<1	$113,562	.150	24	$113,630	.134	27
1-1.99	141,398	.323	54	164,279	.368	52
2-2.99	233,426	.369	47	237,411	.338	34
3-3.99	571,757	.726	37	491,938	.556	40
4-4.99	812,712	.696	44	548,506	.619	28
5-5.99	812,570	.728	32	1,152,236	.523	24
6-6.99	927,118	.595	24	1,079,196	.610	17
7-7.99	876,575	.516	29	920,000	.489	21
8-8.99	1,096,471	.525	17	1,021,067	.485	15
9-9.99	1,008,212	.525	16	1,041,944	.631	9
10-10.99	1,126,021	.549	10	1,030,555	.585	6
11-11.99	630,047	.502	9	979,762	.548	7

of service are presented in Table 5. A cursory inspection of these coefficients indicates that the relative dispersion of salaries nearly doubles after three years of service.

These coefficients of variation support the interpretation that the salaries of arbitration-ineligible players are more tightly structured on the basis of years of service than the salaries of the arbitration-eligible players. Therefore, it is not surprising that above average performance goes unrewarded for the first three years of players' careers. (Power hitters appear to be a partial exception, as the portion of their gap explained by the regression coefficients is lowest and the portion of their gap explained by the performance variables is greatest.) The structuring of salaries for the first three years of players' careers may account for the greater exploitation of many of the above average young players in comparison to the average young players.

REFERENCES

1. *The Baseball Encyclopedia*, 8th edition (New York: MacMillan, 1990).

2. Breusch, T. S., and A. R. Pagan, "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47 (September 1979), pp. 1287-1294.

3. Gunderson, Morley, "Male-Female Wage Differentials and Policy Responses," *Journal of Economic Literature*, 27, no. 1 (March 1989), pp. 46-72.

4. Hadley, Lawrence, and Elizabeth Gustafson, "Major League Baseball Salaries: The Impacts of Arbitration and Free Agency," *Journal of Sport Management*, 5, no. 2 (July 1991), pp. 111-127.

5. Hill, James Richard, "The Threat of Free Agency and Exploitation in Professional Baseball: 1976-1979," *Quarterly Review of Economics and Business*, 25, no. 4 (Winter 1985), pp. 68-82.

6. Jennings, Kenneth M., *Balls and Strikes The Money Game in Professional Baseball* (New York City: Praeger, 1990).

7. Miller, James Edward, *The Baseball Business* (Chapel Hill, North Carolina: University of North Carolina Press, 1990).

8. Mincer, Jacob, *Schooling, Experience, and Earnings* (New York: National Bureau of Economic Research, 1974).

9. Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, no. 3 (October 1973), pp. 693-709.

10. Ramanathan, Ramu, *Introductory Econometrics with Applications* (Fort Worth: Dryden Press, 1992).

11. Scoville, James C., "Labor Relations in Sports," in Roger Noll (ed.), *Government and the Sports Business* (Washington, D.C.: Brookings, 1974).

12. Scully, Gerald W., "Pay and Performance in Major League Baseball," *American Economic Review*, 64, no. 6 (December 1974), pp. 915-930.

13. Scully, Gerald W., *The Business of Major League Baseball* (Chicago: University of Chicago Press, 1989).

14. "1990 Major League Payrolls," *The Sporting News* (April 23, 1990), pp. 22-23.

15. *The 1991 American League Red Book* (New York: The American League of Professional Baseball Clubs, 1991).

16. *The 1991 National League Green Book* (New York: The National League of Professional Baseball Clubs, 1991).