Does the Baseball Free Agent Market Properly Value Performance?

David Gassko[1]
Rice University
Department of Economics
May, 2011

Abstract

Neoclassical models of the labor market assume that given perfect information about a worker's productivity, in a competitive market, employers will pay employees their marginal product. The market for free agents in Major League Baseball provides a perfect opportunity to investigate that theory, as a player's production can be easily observed and his statistics translated to wins. I investigate whether teams properly value the various categories of a player's performance—that is, whether teams properly compensate players for various facets of their production as they relate to wins. Employing a unique data set of multi-year free agent contracts from 1999-2009 and ordinary least-squares (OLS) regression, I find that teams have correctly valued most hitter statistics, with the exception of fielding, which was significantly undervalued in that time period. I then examine related questions, confirming, expanding upon, and on one point, contradicting previous research into the economics of baseball.

[1] I would like to thank my advisor James Brown for his invaluable advice and encouragement, and his extraordinary ability to see the forest, the trees, the pine needles, and the acorns in every claim I have made and regression I have run. All mistakes are my own.

Introduction

It is assumed in the neoclassic model of labor markets that a worker's effort and productivity can be observed, and thus that a laborer's wage will be equal to her marginal product. Unfortunately, such conditions are rarely found in the real world; in the vast majority of industries, only a portion of productivity can be directly observed, making it difficult for employers to appropriately compensate employees and for economists to study compensation. Productivity is thus often considered partially or completely unobservable in Labor Economic literature. Major League Baseball presents a unique situation in which an employee's performance is directly observable, and marginal productivity can be measured. The question arises, then: In a competitive market where productivity can be observed, does a laborer's wage in fact equal her marginal product?

In baseball, a player's productivity can be defined as his contribution to a team's winning percentage.[2] On offense, players generate wins by compiling positive contributions (singles, doubles, triples, etc.) while avoiding negative plays (namely, outs); defensively, the goal is the reverse. A player's value can then be determined by analyzing his numbers in each statistical category. In this paper, I will investigate whether teams appropriately value those contributions relative to their actual value. Using data from 1999-2009, I will estimate a salary equation and compare the estimated coefficients to what a statistical analysis of baseball determines to be the appropriate value for each category.

---

[2] If a team's revenue is dependent on performance-related variables other than winning, the appropriate definition of a player's productivity may be his contribution to a team's revenues, if baseball owners look to maximize profits rather than wins. (Likely, an owner's utility function is in fact a combination of both.) However, it is a common assumption in the literature that revenues are a function of winning, in which case the owner's utility function is of no importance.

The focus of this paper will be on contracts signed in free agency rather than those agreed to in arbitration. Arbitration eligible players are paid less than free agents (Burgers and Walters, 2009), and arbitration precludes a player's salary from being determined in a competitive market. In free agency, on the other hand, 30 teams can bid for a player's services. Free agent contracts are therefore awarded in competitive markets with close to perfect information, making for a theoretically appealing data set.

Teams bid for players in free agency both to add wins, but also to fill out their roster. Baseball is a highly profitable business: Forbes reports that in 2009 the average MLB team had a profit of $17.4 million, with only two teams losing money. Moreover, as Fort (2006) points out, the monetary value of baseball ownership goes beyond operating profit, and includes shelters from federal income taxes, spillovers to other wealth generating elements of the owner's portfolio, and profit taking from the expense side. Additionally, due to a strong revenue sharing system in which 31% of all local revenues are split evenly between the teams as well as all national revenue, it is not necessary for an MLB team to win to be profitable. In fact, according to Forbes, the third most profitable team in 2009 was the Washington Nationals, who lost 103 games.

Therefore, unlike other businesses, baseball teams have a strong incentive to employ a full roster of players regardless of their marginal product to avoid censure from the other teams and to lay claim to what are practically guaranteed profits. This unique feature makes the market for MLB free agents different from other labor markets in an important way; namely, the alternative to signing a given player is signing a minimum-salaried player, rather than no one at all. The correct comparison point for a player's production, then, is the production a team would receive from a minimum-salaried player, rather than no player at all. Prior to estimating a salary equation, I will therefore estimate the expected contribution of a minimum-salaried free agent.

The paper will proceed as follows: Section II will provide a literature review. Section III discusses how runs are scored. Section IV explains the concept of freely available talent. Section V presents my methodology and analysis. Section VI answers some related questions and ties up some loose ends. Section VII concludes and provides directions for future research.

Literature Review

Baseball has long provided a fruitful ground of study for economists. Unlike other industries, performance is easily observable in baseball, making it possible for economists to estimate with some precision a player's marginal product. Moreover, there exists a detailed and accurate data record for Major League Baseball going back to 1871, which is publicly available and easily accessible. Until an arbitrator's ruling struck it down in 1975, MLB contracts contained a reserve clause which tied players to their teams, giving owners total determination over player salaries. The arbitrator's ruling, however, paved the way for free agency, in which all MLB teams are free to bid for a player's services. Thus, for the past 35 years, free agent salaries have been determined through a competitive market process, making them all the more useful for analysis under an economic framework.

The seminal paper in this field is Scully (1974). Scully examines the economic effects of the reserve clause by comparing player salaries with his estimates of their marginal revenue products (MRP). To do so, Scully first estimates an equation relating a team's offensive performance, as represented by slugging average, and its pitching, as represented by strikeout-to-walk ratio, to winning percentage. Scully then relates winning percentage to revenue; joining the two equations allows him to estimate MRP for individual players. He concludes that over their careers, average players under the reserve clause receive about 20 percent of their MRP, while stars receive just 15 percent.

Player salaries have skyrocketed since the abolition of the reserve clause, bringing player compensation more in-line with MRP. Players, however, do not become free agents immediately upon reaching the major leagues. Instead, they play three years under team-determined salaries (essentially equivalent to the reserve clause), and then go through three years of final arbitration before finally gaining the right to free agency. MacDonald and Reynolds (1994) examine data from 1986-87, and find that experienced players (those who have reached free agency) are paid in accord with the estimated MRP, while young players tend to be paid less. Burger and Walters (2009) estimate that players receive between 44 and 64 percent of their MRP in arbitration, with the number increasing as they grow more experienced and near free agency. It is for this reason that I choose to focus on the free agent market for players, while avoiding contracts agreed to outside of free agency.

Many if not most papers in this field try to estimate an equation relating player performance to salary. The complexity of these equations can vary significantly: Scully (1974) uses just one variable for hitters, while Meltzer (2005) includes 14. An important consideration in estimating such an equation is determining which non-performance variables need to be controlled for. Some economists use a team’s payroll rank or market size, but Brown and Jepsen (2009) show that higher revenue teams do not pay more for the same performance. This result is intuitive if one considers the baseball labor market not as a market for players, but as a market for wins, which are provided in different amounts by different players. There is a fixed supply of wins on the free agent market in any given year, and a demand curve based on teams' willingness to pay. There exists then a market clearing price for wins, which all teams will pay.

One consideration in estimating a salary equation is that salary on the free agent market is jointly determined with contract length. Risk-averse players may be willing to trade off less salary for more years. Meltzer (2005) uses a two-stage least squares estimation technique to derive the impact of contract length on salary. He finds that even when controlling for other variables, longer contracts are associated with higher salaries, but that young improving players tend to receive long contracts at low salaries while chronically injured players tend to receive shorter contracts at salaries that are in-line with their performance. Krautmann and Oppenheimer (2002) also use a two-stage least squares technique, but they include a cross term for performance and contract length. Like Meltzer, they also find that longer contracts are associated with higher salaries, but the coefficient on the cross term is negative, implying lower returns to performance as contract length increases.

While this paper will deal with hitters of note is Bradbury (2007), which examines whether the market properly values pitchers based on the statistics they can and cannot control. Bradbury concludes that it mostly does, properly valuing strikeouts and walks relative to their values in preventing runs while slightly overvaluing home runs. Bradbury's article shows rationality in labor market decision on the part of baseball teams, but many other studies find the opposite. Hakes and Sauer (2006) find that teams undervalued on-base percentage until the publication of the book *Moneyball* in 2003.

Burger and Walters (2009) find that teams overvalue pitchers relative to position players in the MLB draft and fail to decrease bonuses adequately for players drafted after the first round. Thus, the question of whether major league teams make rational decisions in the labor market is still an open one, and something I hope to shine some light on in this paper.

How Runs are Scored

The purpose of this paper is to explore whether Major League Baseball teams correctly pay for hitter contributions relative to their actual value, which can be determined empirically. The goal in baseball is to win, and teams win by scoring more runs than their opponents. A hitter's value, then, can be defined as his contribution to the runs scored by his team. The question is how individual statistics can be tied to run scoring on a team level. Baseball statisticians generally use a method known as linear weights, which was introduced by Thorn and Palmer (1984).

Thorn and Palmer observe that the run potential of any given situation in baseball can be categorized based on the bases occupied and the number of outs in the inning. There exist eight potential combinations of bases occupied—no one on, man on first, man on second, man on third, men on first and second, men on first and third, men on second and third, and bases loaded—and three potential numbers of outs—zero, one, or two—as the third out ends the inning. In total, then, there are 24 potential combinations of bases and outs, and we can empirically determine how many runs score on average from each base-out situation to the end of an inning. This number Thorn and Palmer call the run expectancy of the situation; that is, if we there are so many outs in an inning and the following bases are occupied, how many runs can we expect to score before the inning is over. Figure (1) presents a run expectancy matrix for 2008.[3]

Figure 1. Run Expectancy Matrix, 2008

Men on Base	Outs: 0	Outs: 1	Outs: 2
---	0.52	0.28	0.11
1--	0.90	0.53	0.23
-2-	1.15	0.69	0.33
--3	1.50	0.97	0.35
12-	1.53	0.92	0.46
1-3	1.77	1.16	0.48
-23	2.01	1.42	0.59
123	2.31	1.59	0.80

With zero outs and no one on, the run expectancy is 0.52 runs. Not coincidentally, the average major league in 2008 scored exactly 0.52 runs per inning. At the start of an inning, we would obviously expect a team to score 0.52 runs, and then of course it should hold that this would be the case at any point when there are no outs and no men on-base. If the hitter leads off with a single, we now have a runner on first and no outs, resulting in a run expectancy of 0.90 runs. It might be said, then, that the single is worth 0.90 – 0.52 = 0.38 runs in this situation, as it has increased the number of runs we expect the team to score in the inning by 0.38. If the next hitter knocks a home run out of the park, the run expectancy drops back down to 0.52, but two runs also score, so in total it can be said that the run home run was worth 0.52 – 0.90 + 2.00 = 1.62 runs. This exercise can be performed for every single, every home run, and indeed every play that happens in a

[3] Numbers courtesy of Baseball Prospectus (http://www.baseballprospectus.com/).

major league season. Averaging out the change in run expectancy for each event can tell $us the value of that play. This idea Thorn and Palmer call linear weights.

The weights I use are based on Ruane (2005), who calculates linear weight values for each league from 1960-2004. The values vary with run environment: Some categories, such as sacrifice flies, move strongly depending on the number of runs that is scored per game, while others, such as home runs, barely move at all. For each category, I run an ordinary least squares (OLS) regression relating the value of the event to run environment, and then apply those values to the 2008 American League. These are the linear weight values I will use throughout this paper as a benchmark for the correct value of each event. They are presented in Figure (2).

Figure 2. Linear Weights, 2008 American League

Event	**Value**
1B	0.470
2B	0.775
3B	1.057
HR	1.401
BB	0.317
HBP	0.346
OUT	-0.252
SB	0.188
CS	-0.447

On average, a single raises a team’s run expectancy by 0.47 runs, while a home run increases its run expectancy by 1.40 runs. Given this, teams should be willing to pay roughly three times more for each marginal home run than for each single. Any other relationship would imply a mis-valuation of one statistic or the other by the Major

League Baseball market, despite the fact that the linear weights system has been around for 25 years and the values of each event are generally well-known and easily obtainable.

Freely Available Talent

It is imperative for this study to first define what constitutes a player's marginal product. Generally, economists have either used simple counting statistics (such as the number of runs scored by a player, or his totals in some of the categories that will be used in this paper) or a combination or rate statistics (such as the percentage of times a player gets on-base, or the number of bases he gains per at-bat) and playing time statistics (such as plate appearances). I argue that neither approach is correct. Rather, a player's marginal product is his contribution over and above what a free agent available for the Major League minimum salary would produce.

The economics of Major League Baseball ensure that every team will want to fill its roster to the 25-man limit set by the MLB. A team that did not employ players at every position would be put at a significant competitive disadvantage, one which all but ensures that the marginal revenue product (MRP) of every player as compared to employing no one at all will be greater than zero. Moreover, a team that did not fill all of its roster spots would likely face extraordinary criticism from both its fans and other teams, the former significantly impacting local revenues and the latter creating the risk of the owners losing their franchise rights, which are extraordinarily valuable. Since a team that employs minimum-salaried players at all positions is still guaranteed a hefty profit—teams appear to receive roughly $40 million apiece from the MLB Central Fund according to Deadspin (2010), whereas a payroll consisting of only minimum-salaried players would be just $10 million—it is inevitable that all teams will do so. It only takes a look at off-season chatter by fans and columnists to confirm this fact: The question is never *if* the hometown team will sign someone to play shortstop, but *who*.

Of course, we can test to see whether Major League teams in fact value a player's marginal product as his performance over and above that of a minimum-salaried player rather than above zero, and I will do so later in this paper. However, I am forced to first calculate and use the minimum salary baseline so that I can build a performance metric, which will be important for calculating baseball-specific inflation over time.[4] Calculating the minimum-salary baseline is a relatively simple task. I take all players from 1985-2009 who signed as free agents in the off-season for no more than twice the minimum salary in that year[5] and adjust their numbers to the context of the 2008 American League (a necessary step, as offensive levels in Major League Baseball are highly inconsistent from year to year). The next step could be as simple as summing all those numbers and calculating the average for each position, however the results of such an operation would be biased upward and unrealistic. This is a problem that Silver (2006) ran into when he attempted a similar exercise, and I will quote his explanation for its lucidity:

> Say that the Mariners' Pacific Rim scout is on a flight with a pilot that has had far too much sake and happens to land in Yakutsk, Sibera instead of Sapporo, Japan. The intrepid scout can't get a flight back until the next day and decides to take in a Yakutsk Yaks game, where he discovers a set of twin left-handed pitchers, Miroslav Borscht and Radoslav Borscht, who each hit 95 on his JUGS gun. The twins are given non-roster invitations, and an all-expenses paid trip to

[4] As I will show in the next section, baseball salaries have increased at a much faster pace than the general rate of inflation over the past decade, and this is something that must be adjusted for.
[5] I do not limit myself to players who were paid strictly the minimum so as to significantly expand the sample size. The results if using minimum-salaried players would be similar, but less stable when broken down by position. Twice the minimum salary is still chump change in baseball terms.

> Peoria, Arizona. Miroslav turns out to be the Ozzie Canseco of the pair, with a weakness for flavored vodkas and Maricopa County's finest topless bars, and is cut a week into camp, while Radoslav emerges as the team's second starter.
>
> Obviously, we'd have a selective sampling issue if we gave full credit to Radoslav's performance, while forgetting about Miroslav entirely—one of the risks when signing an unknown player is that you may waste a significant number of at bats on him before you figure out that he's not even qualified to carry Mario Mendoza's jock. The chosen solution was to "max" everyone's playing time at the rookie minimum of 130 PA. So, even if the player played a full season, his statistics were weighted as though he'd only had 130 PA. If the player had fewer than 130 PA, then his playing time was taken as is.

I employ Silver's solution, capping each player's statistics at a weight of no more than 130 plate appearances. Figure (3) presents my results, pro-rated to roughly a full season of plate appearances (700).[6]

Figure 3. Performance by Freely Available Talent (FAT), by Position

Position	1B	2B	3B	HR	SB	CS	BB	HBP	Park	BSrun	Field	LW
C	99	30	1	13	1	1	56	7	1	-4	-2	-39
1B/DH	104	32	2	17	4	2	69	6	-1	-2	-1	-15
2B	116	32	3	10	9	3	58	7	0	1	-5	-20
3B	102	31	3	13	6	3	61	6	-1	0	-2	-27
SS	110	30	3	9	6	3	46	5	0	-1	-6	-41
LF	105	32	4	12	13	4	67	5	-1	2	0	-13
CF	111	27	4	12	22	9	60	6	1	0	0	-23
RF	106	32	3	19	7	2	58	7	1	0	-1	-13

[6] Note that some less important categories have been left out of the table for space considerations, but are included in the linear weight calculation and are available from the author upon request.

The results confer with intuition: Minimum-salaried players, or Freely Available Talent (FAT), as I will call the baseline from now on, are well below-average, though just how far below varies, with corner outfielders clocking in at -13 runs while FAT shortstops are an incredible -41 runs per 700 PA. That makes sense: Shortstop requires much greater defensive skill, and so a minimum-salaried shortstop will not be nearly as good a hitter as a minimum-salaried left fielder. We also see that the skill composition at each position varies: Right fielders and first basemen tend to hit twice as many home runs as shortstops, while catchers tend to be significantly below average base runners.

Methodology & Analysis

From a theoretical standpoint, I assume that team owners maximize their own utility, which is a function of profit and winning. Profit, of course, is simply revenue minus costs, or formally,

$$\pi = R - C$$

Revenues include a rent, α, that teams receive simply for playing (which includes revenue received from the MLB central fund, plus revenues from attendance and local media contracts that will accrue to a team filled with minimum-salaried players), plus a variable amount that is dependent on the team's marginal win total:

$$R = \alpha + r(W)$$

Costs, meanwhile, include a fixed cost of operations (including the cost of a roster of minimum-salaried players), φ, plus a variable cost that is dependent on the number of marginal wins that the team purchases:

$$C = \varphi + c(W)$$

Combining the two functions, we get:

$$\pi(W) = \alpha - \varphi + r(W) - c(W)$$

Since α and φ are constants, we can simply observe that so long as $\alpha > \varphi$, teams will always choose to incur the fixed costs of operations, and thus always at least hire a roster full of minimum-salaried players. If owners derive direct utility from each win,

$$U = U(\pi(W), W)$$

Maximization of utility will imply that,

$$\frac{\partial U}{\partial \pi} \times \frac{d\pi}{dW} + \frac{\partial U}{\partial W} = 0$$

Re-arranging the equation, we get:

$$\frac{d\pi}{dW} = -\frac{\frac{\partial U}{\partial W}}{\frac{\partial U}{\partial \pi}}$$

Assuming that an owner derives positive utility from winning (that is, $\partial U/\partial W > 0$), the implication is that s/he will spend more to purchase a marginal win than that win is worth in terms of revenue. Each owner's utility function is of course unique; some owners surely value winning more than others, and some teams surely derive more revenue from each marginal win than do others. At the same time, however, all owners ultimately purchase wins in one market, and so the cost of a win is dictated by the law of one price. At the margin, each owner will value wins identically; it just takes a different number of wins for each owner to arrive at that margin. The cost of a marginal win in the free agent market, then, is determined by the intersection of the supply of wins and the sum of the owners' demand curves.

For the empirical test, I employ a unique database of free agents signed between off-seasons of 1999-2009 for the empirical analysis in this paper.[7] Data for 2008-2009 was taken from the ESPN Free Agent Tracker, while data for the previous seasons was generously donated by Dave Studenmund of The Hardball Times. I limit myself to players who signed multi-year contracts, as players who sign one-year deals tend to have significantly different characteristics and also can receive a significant portion of their compensation through incentives and performance bonuses which are not captured in my database. That leaves 240 players in my sample, with contract lengths ranging from 2-10 years, and salaries from $300,000-$27,500,000 per year.

Contract amounts must be adjusted for inflation, but doing so is not straightforward. Multi-year contracts are signed without knowledge of future inflation, so it is necessary to use some form of inflation expectations. While it would probably be optimal to use TIPS market spreads at the time of the contract signing, I employ a simpler solution, observing that inflation consistently averaged 2.5% over the period covered by my data sample. I therefore assume a stable inflation expectation of 2.5% per year, and convert all contracts into 2009 dollars. Contracts are treated as an annuity, an important feature since longer-term contracts will tend to have a lower present value than shorter deals signed for the same dollar-per-year amount. I therefore use the following equation to estimate an adjusted per-year contract amount, where *Length* is the length of the contract, *Guarantee* is the amount of salary guaranteed throughout the contract, and *Year* is the year the contract was signed:

[7] The years signify the first year of a player's contract, meaning that the earliest free agents in my sample first played under their new contracts in 1999, while the last free agents first played under their new contracts in 2009.

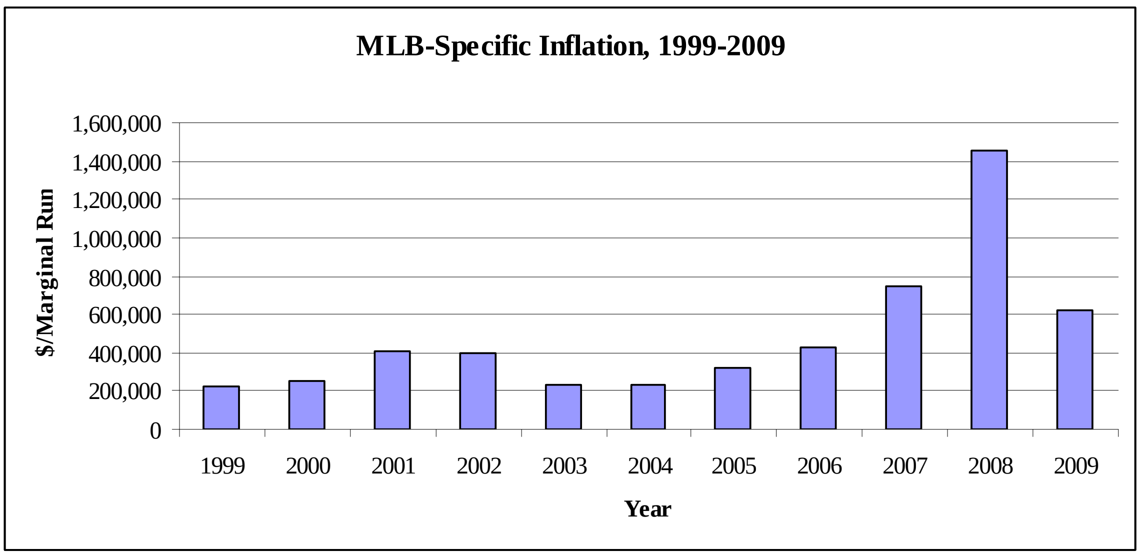
$$\textit{Adjusted \$/Year} = \frac{1-\dfrac{\left(\dfrac{1}{1.025}\right)^{Length}}{.025}}{Length^{2}} \times Guarantee \times 1.025^{(2009-Year)}$$

One potential pitfall of this method is that salaries are assumed to be constant throughout the time of the contract. This oftentimes is not true—contracts tend to be back-loaded, with larger salaries being paid out in the latter years—but the real problem occurs if the bias is not consistent; that is, if the term structures of contracts vary. Insofar as they do, the adjusted salary amounts I use may be incorrect, biasing the final results. Unfortunately, I could not locate an easily accessible source containing term structure information for the years studied, so we must hope that any existent bias is small.

It is not, however, enough to adjust baseball salaries for economy-wide inflation. One thing previous empirical examinations of the MLB labor market have not done is also adjust for baseball-specific inflation. This omission severely biases the results, as contracts signed in later years tend to look much greater than those that came before them. For example, in 1999, Roberto Alomar signed a contract that paid him $8 million per year. He was an All-Star each of the next three years, and a top-four MVP candidate two of them. In 2007, Nomar Garciaparra also signed a deal paying him $8 million a year. He proceeded to contribute minimal to no value over a FAT player over the next three years of his career. Clearly, by 2007, $8 million didn't buy what it used to.

I compute the marginal product for each free agent in my sample by calculating his statistics in the first year of his contract (adjusted to the context of the 2008 American League) relative to a FAT player at his position, and then applying to those numbers the linear weights derived earlier in this paper. I then calculate his marginal salary—that is, the salary he was paid above the league minimum—adjusted for economy-wide inflation, and find the number of dollars in each year that were required to purchase a marginal run on the free agent market. As Figure (4) clearly shows, though that number varies substantially from year-to-year due to random variation, the overall trend is clearly inflationary. Overall, the price of a marginal run increased from $225,336 in 1999 to $619,633 in 2009, an annualized inflation rate of 10.6%.[8] Therefore, I adjust the figure derived from Equation (1) by a further 10.6% per year to express all salaries in 2009 MLB-equivalent dollars.

Figure 4. MLB Inflation-Adjusted $/Marginal Run, 1999-2009



Just as I adjust salaries for inflation, I also adjust all players' numbers to put them into the context of the 2008 American League, as offensive levels have fluctuated significantly over the years of my sample. Figure (5) shows the statistical categories I use, as well as their means and standard errors within the sample. I use a player's statistics in the first year of his contract, implying rational expectations on the parts of

[8] It is important to remember that this number is calculated after adjusting all contracts for economy-wide inflation. Therefore it can be said overall free agent salaries over the past decade have increased by about 10.6% + 2.5% ≈ 13% per year.

Major League Baseball teams. Though it would be preferable to use a player's performance over the length of his contract, that would severely cut down the sample by excluding players whose contracts have not yet ended. In addition to traditional hitting categories, I include the number of runs above or below average that the player contributed on the bases, the number of runs above or below average that he saved in the field, and run impact of his ball park so that players who sign with teams that play in good hitters' parks do not look better than they actually are and vice-versa. Those numbers are then compared to what a FAT player would have contributed in the same number of plate appearances to arrive at the player's marginal contribution in each category. The vector of those marginal contributions is the player's marginal product. Since a marginal run cost $619,633 in 2009, we would expect a player's marginal salary (that is, his salary minus the league minimum) divided by that number to be equal to:

$$\frac{mSal}{619{,}633} = \begin{array}{l} 0.470\times m1B + .775\times m2B + 1.057\times m3B + 1.401\times mHR + .317\times mBB \\ +.346\times mHBP + .188\times mSB - .447\times mCS - 1.000\times Park \\ +1.000\times mBSrun + 1.000\times mField \end{array}$$

I empirically estimate this equation, controlling also for whether or not a player was re-signed. The estimated equation takes the following form, where *X* is the vector of production variables, *R* is a dummy term indicating whether or not the player was re-signed by his team, and ε is the error term:

$$\frac{mSal}{619{,}633} = \beta X + \lambda R + \varepsilon$$

The results of the regression are shown in Figure (6). If baseball teams are correctly valuing hitters, then the empirically estimated coefficients should match the expected coefficients presented above. Thus, I compare the estimated and actual coefficients, and calculate whether or not any differences are statistically significant.

Figure 5. Variable Definitions and Summary Statistics

Variable	Definition	Mean	S.E.
1B	Singles.	71.2	33.726
2B	Doubles	21.7	10.926
3B	Triples.	1.9	2.178
HR	Home runs.	12.4	10.842
BB	Walks.	42.5	27.914
HBP	Hit-by-pitch.	3.9	3.425
SB	Stolen bases.	6.4	8.938
CS	Caught stealing.	2.3	2.502
Park	The run impact of the player's park, defined as the number of extra runs the player would be expected to create due to his ballpark.	0.0	2.314
BSrun	Base running runs, as calculated by Smith (2010).	0.1	2.211
Field	Fielding runs, as calculated by Smith (2010).	-0.2	7.910

Figure 6. Regression Results

Variable	β	S.E.	Expected	T-Value	P-Value
Re-Sign	4.633	1.684	0.000	2.75	0.006
m1B	0.565	0.072	0.470	1.32	0.188
m2B	0.736	0.160	0.775	-0.24	0.808
m3B	-0.300	0.486	1.057	-2.79	0.006
mHR	1.327	0.121	1.401	-0.61	0.541
mSB	0.283	0.197	0.188	0.48	0.630
mCS	-0.315	0.597	-0.447	0.22	0.825
mBB	0.363	0.055	0.317	0.84	0.404
mHBP	0.094	0.295	0.346	-0.85	0.394
run_imp	-0.641	0.368	-1.000	0.98	0.330
mBSrun	0.684	0.546	1.000	-0.58	0.563
mfielding	0.018	0.104	1.000	-9.44	0.000

R^2 = 0.709
F = 46.38
S.E. = 12.84

Only three categories show significant results: Re-signing, triples, and fielding. Teams appear to pay more for players they re-sign, indicating that there is no such thing as a hometown discount as far as the MLB labor market is concerned; in fact, the effect is very much the opposite, with teams paying more (by a factor of almost $3 million per year) to re-sign their own players. This might be a rational response to the fact that teams know more about their own players, and thus might out-bid competitors for hitters that they expect to play well, while letting go those that they expect to decline. Indeed, Swartz (2010) finds evidence that re-signed players perform better than those who leave for another club. The negative coefficient on triples is a mystery, but not a particularly important one. As shown in Figure (5), the average hitter in the sample averaged fewer than two triples per season, with a relatively tight spread at that. Moreover, since triples likely decline strongly with aging due to their great dependence on a player's speed, it is likely that both the mean and spread decline further past the first season of a player's contract. Given these facts, it is clear that triples make up a very small proportion of a hitter's value, and while that is no reason for MLB teams to ignore them, it is an indication that such ignorance is of relatively little importance.

The final significant coefficient, however, does seem quite important. The regression finds that teams seem to place absolutely no weight on a player's fielding abilities, even though they make up a large portion the average player's marginal product. Returning to the summary statistics in Figure (5), it is apparent that only singles and home runs have a significantly greater spread in terms of runs than does fielding, implying that only those two categories should be of clearly greater importance to Major League general managers. Instead, from 1999-2009, MLB teams appeared to place no weight on a player's fielding ability when offering him a contract. Anecdotal evidence serves as confirmation of the empirical findings. In 2010, a multitude of media stories focused on a new emphasis being placed by teams on fielding and statistical measures of fielding ability. See, for example, Benjamin (2010), discussing the fielding-focused off-season acquisitions by the Boston Red Sox, as well as similar strategies adopted in recent years by the Tampa Bay Rays and Seattle Mariners, and Gregory (2010), where the author posits that "a sharper focus on defense among front offices has clearly assisted" in lowering run scoring across the MLB. Benjamin (2010) also provides an insightful quote from Red Sox General Manager Theo Epstein: "Teams always understood the value of defense," Epstein explains. "I think, if anything, the explosion of offense in a certain period 10 to 15 years ago might have swung the pendulum a little too far in the other direction. *It was probably being overlooked too much.*" [Italics added.] Epstein's anecdotal impressions are confirmed by the empirical results, and vice-versa. It appears that even in the competitive baseball market for labor, inefficiencies can still exist.

Related Questions

Did Moneyball *change how teams value players?*

Michael Lewis' best-seller *Moneyball*, released in 2003, chronicled the unparalleled success of the Oakland Athletics, who managed to win 481 games in the five years prior to the book's publication, an average of 96 wins per season, while operating on a consistently small budget. Lewis attributed the Athletics' success to their ability to find under-valued players, specifically players with high on-base percentages but otherwise unimpressive numbers. On-base percentage (OBP) measures how often a player makes it on base (or, to put it another way, avoids making an out) on a per plate appearance basis. Until the publication of *Moneyball*, on-base percentage was a fairly obscure statistic, shunned in favor of the more traditional batting average (BA), which measures a hitter's propensity to get a hit in a given at-bat. The essential difference between BA and OBP is that OBP adds walks and hit-by-pitch both to the numerator and denominator employed in the batting average calculation (it also adds sacrifice flies to the denominator, but that's a small factor). In other words, Lewis' argument could be re-stated to say that teams, at least from the years 1999-2003 were undervaluing walks (and, to a lesser extent, hit-by-pitch, which are simply a much less common event with a much smaller spread in the population—see Figure 5).

Hakes and Sauer (2006) purports to find strong evidence for Lewis's thesis. They estimate earnings equations for each of the seasons from 2000-2004 and find that OBP was significantly undervalued until the publication of *Moneyball*, but that in the season after the book's publication, 2004, the labor market corrected, and OBP became properly valued thereafter. Hakes and Sauer (2007) expands the study to a greater number of years, and re-affirms the conclusions reached in the earlier paper. There seem to be, however, a number of problems with the two papers. First of all, both papers include in their samples all players with at least 130 at-bats in the previous season. Not only does that mean that the samples used in those papers include a significant number of players whose salaries were not determined in the free market (players are eligible to become free agents only after playing for six years in the major leagues), but it also means that most players included in their samples did not in fact have their salary determined prior to the season examined. For example, many if not most players in the 2004 sample, which Hakes and Sauer (2006) argues shows a drastic change in player valuation, in fact signed their contracts prior to the publication of *Moneyball*, so that a dramatic shift in the OBP coefficient from "incorrect" to "correct" could have occurred only due to statistical variation, or, in fact, a dramatic *overcorrection* of the market. This is one problem; another is that Hakes and Sauer (2006) defines performance variables (on-base percentage and slugging average), positional variables (catcher and infielder), and playing time variables (plate appearances) separately, whereas, as should be clear from the section of this paper defining Freely Available Talent, these variables all interact and should be treated multiplicatively. Teams pay a catcher to play consistently and to hit; they do not pay someone to be a catcher, and then additionally to play, and then additionally to hit. As such, Hakes and Sauer's independent variables suffer from misspecification and collinearity.

The question raised by Hakes and Sauer (2006), however, is one of interest: Did the labor market undervalue walks (or hit-by-pitch) until the publication of *Moneyball*, and did it correct after 2003? Such a question can easily be answered within the context of this paper, merely by splitting the sample used in two, and running the regression seen in Figure 6 separately for the years 1999-2003 and 2004-2009. The results of these regressions can be seen in Figures 7 and 8 below. I also run a broader regression, which uses dummy variables interacted with the performance variables, multiplying by 1 for seasons from 1999-2003 and by 0 for 2004-2009. The results for that regression are printed in Figure 9.

Figure 7. Regression Results, 1999-2003

Variable	β	S.E.	Expected	T-Value	P-Value
Re-Sign	-0.078	3.425	0.000	-0.023	0.982
m1B	0.789	0.147	0.470	2.17	0.032
m2B	0.611	0.301	0.775	-0.54	0.589
m3B	-0.111	0.900	1.057	-1.30	0.198
mHR	1.858	0.236	1.401	1.94	0.056
mSB	0.275	0.395	0.188	0.22	0.825
mCS	0.310	1.030	-0.447	0.73	0.465
mBB	0.380	0.099	0.317	0.63	0.528
mHBP	0.805	0.614	0.346	0.75	0.457
run_imp	-0.964	0.621	-1.000	0.06	0.954
mBSrun	0.686	1.123	1.000	-0.28	0.781
mfielding	0.077	0.191	1.000	-4.83	0.000

$R^2 = 0.780$
$F = 24.88$
$S.E. = 14.93$

Figure 8. Regression Results, 2004-2009

Variable	β	S.E.	Expected	T-Value	P-Value
Re-Sign	7.458	1.643	0.000	4.54	0.000
m1B	0.411	0.075	0.470	-0.79	0.430
m2B	0.979	0.173	0.775	1.17	0.242
m3B	-0.050	0.513	1.057	-2.16	0.033
mHR	0.903	0.126	1.401	-3.94	0.000
mSB	0.398	0.205	0.188	1.03	0.306
mCS	-0.743	0.670	-0.447	-0.44	0.659
mBB	0.248	0.068	0.317	-1.01	0.315
mHBP	-0.299	0.285	0.346	-2.26	0.025
run_imp	-0.579	0.455	-1.000	0.93	0.356
mBSrun	0.369	0.587	1.000	-1.08	0.284
mfielding	-0.027	0.107	1.000	-9.64	0.000

$R^2 = 0.702$
$F = 28.18$
$S.E. = 9.98$

Figure 9. Regression Results, 1999-2009 with Dummy Variables

Variable	**β**	**S.E.**	**Expected**	**T-Value**	**P-Value**
Re-Sign	7.458	1.999	0.173	3.730	0.000
m1B	0.411	0.091	0.233	4.504	0.000
m2B	0.979	0.211	0.242	4.639	0.000
m3B	-0.050	0.624	-0.004	-0.080	0.936
mHR	0.903	0.154	0.363	5.871	0.000
mSB	0.398	0.249	0.126	1.598	0.111
mCS	-0.743	0.815	-0.065	-0.912	0.363
mBB	0.248	0.083	0.196	2.983	0.003
mHBP	-0.299	0.347	-0.039	-0.863	0.389
run_imp	-0.579	0.553	-0.058	-1.046	0.297
mBSrun	0.369	0.714	0.034	0.516	0.606
mfielding	-0.027	0.130	-0.010	-0.211	0.833
Re-Sign	-7.536	3.430	-0.107	-2.197	0.029
m1B	0.378	0.150	0.127	2.518	0.013
m2B	-0.367	0.323	-0.057	-1.135	0.258
m3B	-0.061	0.962	-0.003	-0.063	0.950
mHR	0.955	0.246	0.258	3.883	0.000
mSB	-0.123	0.406	-0.023	-0.302	0.763
mCS	1.054	1.169	0.060	0.901	0.369
mBB	0.132	0.116	0.078	1.136	0.257
mHBP	1.104	0.608	0.081	1.816	0.071
run_imp	-0.385	0.749	-0.030	-0.514	0.608
mBSrun	0.318	1.160	0.020	0.274	0.784
mfielding	0.105	0.203	0.024	0.516	0.606

R^2 = 0.754
F = 27.55
S.E. = 12.15

The results in the first two regressions overall correspond well with the earlier findings in this paper, the one exception being that home runs appear to have been undervalued in the years 2004-2009. Without a valid explanation for that finding, however, it seems safest to assume that the result, though statistically significant, is merely a product of random noise, which will cause even probabilistically unlikely results to pop up given enough variables and tests. As for walks, however, they do not appear to have been undervalued in the free agent market prior to the publication of *Moneyball*, and it does not appear that teams marked up the value of walks thereafter. In fact, though both results are not significantly different from the expected coefficient of 0.317, the observed coefficient on walks actually drops from 1999-2003 to 2004-2009, from 0.380 to 0.248.

The regression in Figure 9 shows similar results to what we find in the separate regressions. Notably, it does provide a better overall fit than the regression in Figure 6 ($F = 3.24$, $p < 0.001$), indicating that there was a significant difference in how teams valued the variables of player performance *jointly* before and after the publication of *Moneyball*. If anything, however, that change was due to teams placing less weight on singles and more on home runs, rather than a greater appreciation for high-walk hitters.

Do teams place too much weight on recent years?

In much of this paper, I try to identify whether the MLB labor market correctly values various statistical contributions by individual hitters, but there also exists the question of whether they correctly weigh past performance in order to project a player's future marginal product. For example, psychologists often refer to the recency effect, which suggests that human beings tend to recall recent events better than those that came before them. The recency effect might cause teams to place too much weight on recent performance, while discounting what a player has done in prior seasons. Anecdotally, there exists the idea that some players perform better in their "contract year" (the season before their contract ends) in order to obtain a contract that pays greater than the value of their expected future marginal contract. In *Baseball Between the Numbers* (2006), author Dayn Perry finds that players as a whole tend to outperform expectations in their contract years, due primarily to an increase in playing time, lending some credence to the "contract year" theory, and giving cause to believe that players might expect teams to overweight recent results.

Procuring an answer to this question is a straightforward exercise. I apply the value formula derived in Figure 6 (minus the re-signing variable) to the three seasons prior to free agency for each player in my database, producing his marginal revenue product for those three seasons as it might be evaluated by the teams. I then regress those variables against the player's *actual* marginal revenue product in his first season after free agency (again using the coefficients from Figure 6) as well as his salary. The equations are as follows:

$$MRP_t = \alpha + \beta_1 MP_{t-1} + \beta_2 MP_{t-2} + \beta_3 MP_{t-3} + \varepsilon$$

$$mSal = \alpha + \beta_1 MP_{t-1} + \beta_2 MP_{t-2} + \beta_3 MP_{t-3} + \varepsilon$$

If teams are suffering from recency bias, then the coefficient on the last season prior to free agency should be greater when salary is the dependent variable than it is when regressed against actual marginal revenue product. The results for these two regression (first with actual marginal product as the dependent variable, then salary) are shown in Figures 10 and 11.

Figure 10. Regression Results, Actual Marginal Revenue Product

Variable	β	S.E.	T-Value	P-Value
α	7,068,091	656,853	10.761	0.000
MP_{t-1}	0.177	0.040	4.398	0.000
MP_{t-2}	0.103	0.039	2.621	0.009
MP_{t-3}	0.137	0.040	3.422	0.001

$R^2 = 0.282$
$F = 30.97$
$S.E. = 8{,}030{,}205$

Figure 11. Regression Results, Salary

Variable	β	S.E.	T-Value	P-Value
α	4,160,054	688,108	6.046	0.000
MP_{t-1}	0.143	0.042	3.384	0.001
MP_{t-2}	0.053	0.041	1.285	0.200
MP_{t-3}	0.176	0.042	4.176	0.000

$R^2 = 0.228$
$F = 23.20$
$S.E. = 8{,}412{,}304$

The two regressions show fairly similar results, and none of the differences between the coefficients are statistically significant (the *p*-values for the differences between the coefficients on MP_{t-1}, MP_{t-2}, and MP_{t-3} are 0.419, 0.224, and 0.354, respectively). In other words, it appears that teams do not overweight recent results, correctly understanding the value of recent performance metrics relative to more dated numbers. One potential snag that does bear noting, however, is that in both regressions a player’s performance from three years prior appears to be more important than his performance from two years prior. This is surely a statistical fluke—those who project baseball players professionally have concluded that for hitters, each season should receive roughly 25% more weight than the one that came before it.[9] As both regressions show the same effect, it seems likely that it can be attributed to a somewhat flukish sample rather than random noise in the regressions themselves. Thus, further researchers might consider using a larger sample by taking into account arbitration-eligible players, as well as a great number of seasons to re-test these results.

Does performance relate to revenue in the same way as it relates to wins?

The model in this paper assumes that an owner's utility is a function of his team's win total, both as it affects revenue as well as the additional utility derived from owning a winning team, and that therefore owners will select a number of wins that maximizes their utility. Insofar as winning affects revenue, it might be useful to think about how a team's chances of winning affect the attendance it draws for a given game, as attendance makes up by far the largest portion of a team's variable revenue. The crowd that a team, i, draws for a given game, j, can be modeled as a function of the team's market ($Market_i$), the quality or newness of its stadium ($Stadium_i$), its chances of winning that game (Win_{ij}), and the quality of its opponent ($Opponent_j$). Presumably, teams from larger markets will draw larger crowds, as will teams with better or newer stadiums. As a team's chances of winning go up, fans are likely to show up in greater numbers to root them on, and better opponents should too bring in more fans, since the quality of play will be higher. Mathematically, the model can be expressed as such:

$$Attendance_{ij} = f(Market_i, Stadium_i, Win_{ij}, Opponent_j)$$

[9] See Cartwright (2010).

Over a full season, I assume that the quality of opponents will average out, and the team's chances of winning each game will add up to its final record.[10] So, over a full season, the model will add up to the following:

$$Attendance_i = f(Market_i, Stadium_i, Win_i)$$

If owners select a number of wins that maximizes their utility, a rational market would assign the proper relative value to each input of winning—that is, if a home run adds three times as many wins as a single, then teams should be paying three times as much for a home run. In the main section of this paper, I find that this expectation holds empirically with the exception of fielding, which appears to be undervalued in the MLB free agent market. However, if a team's revenues are related to the underlying inputs of winning in a different manner than that in which those inputs contribute to winning, the expectation that is applied throughout this paper would itself be incorrect. For example, Gassko (2008) finds that, controlling for a team's win total (and therefore, their effect on winning) home runs tend to lead to higher attendance. In that case, if singles were not associated with a team's attendance beyond their affect on winning, it would in fact be rational for teams to pay more than three times as much for a home run as for a single, as each home run would result in three times as many wins but also additional profit. In the context of this paper, this provides a rational explanation for why teams might undervalue fielding: Perhaps fans simply don't show up to watch good defense, so while quality fielders contribute extra wins, they provide no additional profits. Conversely, if Gassko (2008) is correct, that could mean that teams are actually undervaluing home runs even though they correctly value them in terms of their effect on wins.

[10] It would be interesting to try to model attendance using game, rather than season, level data. This would certainly be an interesting direction for future research.

This proposition is easy enough to test, and given the importance of this assumption to my model, it certainly is necessary. Though technically the proper dependent variable to test would be a team's variable revenue, I use attendance as a proxy since reliable revenue data is not available. Most of a team's variable revenue is derived from ticket sales, and that which is not (i.e. promotions, parking, and stadium advertising) still should bear a strong correlation, so attendance should make for a good stand-in. I then specify two regressions, the first to show the relationship between wins and attendance, controlling for a few key variables, and the second expanding on the first to include all the underlying variables affecting wins that were tested in the regression in Figure 6. The first equation I specify appears as follows, where *Wins* denotes a team's win total, *Playoffs* is a dummy variable indicating whether or not it made the playoffs, *Expansion* a dummy variable indicating whether or not it was an expansion team, *NewStadium* whether or not the team was playing in a new stadium, *MarketSize* is the team's market size per Silver (2007), with an average market indexed to 100, and the subscripts are the same as throughout the rest of the paper:

$$Att. = \begin{array}{l} \alpha + \beta_1 Wins_t + \beta_2 Wins_{t-1} + \beta_3 Playoffs_{t-1} + \beta_4 Expansion_t + \beta_5 Expansion_{t-1} + \\ \beta_6 NewStadium + \beta_7 MarketSize + \varepsilon \end{array}$$

The second regression replicates the first, but with the vector of performance variables from the regression in Figure 6 added in. I also include two additional, important controls: The number of runs a team was expected to score based on its component statistics (as provided by the BaseRuns formula[11]) and the number of runs it was expected to allow. These controls are pivotal because a team's won/loss record may be affected random luck and variation over the course of a season. If a team scores fewer runs than

[11] See Heipp.

might be expected based on its component statistics (or allows more), its win total will, on average, be adversely affected; however, fans might base their assessment of team quality based on their expected runs scored (allowed) rather than the actual number, as that number is more indicative of the team's actual quality of play. Therefore, if we did not control for expected runs scored and allowed, we might (and indeed would) find that various component statistics are positively correlated with attendance, even controlling for winning and other factors, when really we would be suffering from omitted variable bias. The second regression takes the following form, where *BaseRuns* is the expected runs scored by the team, *BaseRunsA* is the expected runs allowed by the team, and *X* is the vector of production variables:

$$Att. = \begin{array}{l} \alpha + \beta_1 Wins_t + \beta_2 Wins_{t-1} + \beta_3 Playoffs_{t-1} + \beta_4 Expansion_t + \beta_5 Expansion_{t-1} + \\ \beta_6 NewStadium + \beta_7 MarketSize + \beta_8 BaseRuns + \beta_9 BaseRunsA + \lambda R + \varepsilon \end{array}$$

My dataset consists of all team-seasons from 1999-2008 (300 in all), with all variables except for the dummies adjusted for league averages. The results of the two regressions are presented in Figures 12 and 13, respectively.

Figure 12. Regression Results, Wins and Attendance

Variable	**β**	**S.E.**	**T-Value**	**P-Value**
α	-500,699	315,921	-1.58	0.114
$Wins_t$	15,279	3,230	4.73	0.000
$Wins_{t-1}$	16,506	4,036	4.09	0.000
$Playoffs_{t-1}$	115,280	100,168	1.15	0.251
Expansion	402,658	538,418	0.75	0.455
$Expansion_{t-1}$	166,798	312,770	0.53	0.594
NewStadium	612,807	177,503	3.45	0.001
MarketSize	3,408	547	6.22	0.000

R^2 = 0.468
F = 36.77
S.E. = 535,545

Figure 13. Regression Results, Including Performance Variables

Variable	β	S.E.	T-Value	P-Value
α	-77,491	3,579,349	-0.02	0.983
$Wins_t$	1,885	5,210	0.36	0.718
$Wins_{t-1}$	12,549	3,906	3.21	0.001
$Playoffs_{t-1}$	156,347	93,784	1.67	0.097
adj1B	590	3,907	0.15	0.880
adj2B	-9,214	6,161	-1.50	0.136
adj3B	-5,442	9,386	-0.58	0.563
adjHR	-2,426	10,886	-0.22	0.824
adjSB	-1,658	1,966	-0.84	0.400
adjCS	2,368	4,559	0.52	0.604
adjBB	-1,313	2,744	-0.48	0.633
adjHBP	-4,046	3,699	-1.09	0.275
adjField	2,067	921	2.24	0.026
adjBsRun	-4,053	5,063	-0.80	0.424
BaseRuns	6,630	7,750	0.86	0.393
BaseRunsA	-451	587	-0.77	0.443
Expansion	889,042	513,835	1.73	0.085
$Expansion_{t-1}$	139,137	290,554	0.48	0.632
NewStadium	495,292	166,644	2.97	0.003
MarketSize	3,249	544	5.98	0.000

$R^2 = 0.566$
$F = 19.22$
S.E. = 494,147

The results in Figure 13 show that no production variable was significant at $p < 0.01$. In other words, attendance does not appear to be influenced by the makeup of a team's production, outside of how that production influences its win total, validating the theoretical model and this paper's main empirical results. Interestingly, the fielding variable is significant at $p < 0.05$, but more interestingly still, the coefficient on the variable is positive, meaning that good fielding teams appear to draw *more* fans than one would expect based on their won/loss record. Though I inclined to dismiss this as a result of random noise (I am, after all, testing 19 variables in this regression—we should expect one of them to show up significant at $p < 0.05$ simply due to random chance), this does indicate that if anything, MLB teams should be paying *more* for good fielders than their contribution to the team's win total alone would indicate, rather than seemingly ignoring fielding completely in valuing free agents.

Another result to note in Figure 13 is that, in contrast to the regression in Figure 12, the coefficient on the $Wins_t$ variable drops from 15,279 to 1,885 and goes from being highly significant ($t = 4.73$) to highly insignificant ($t = 0.36$). The reason for this is the inclusion of the *BaseRuns* and *BaseRunsA* variables, which are better correlated with team quality than wins themselves, as they are directly tied to performance, whereas wins are a combination of a team's performance and the random variation that affects the timing of that performance.[12] Since attendance is a function of a team's expected chances of winning (see the theoretical model at the beginning of this section), and its odds of winning are dependent on the team's expected performance, it stands to reason that attendance can be better predicted using variables that are more directly tied to a team's performance. The regression in Figure 13 appears to confirm that supposition. In fact, if I altogether remove wins from the regression, both the F-statistic and the adjusted R^2 improve, from 19.22 and 0.537 to 20.35 and 0.538, respectively. The results for this regression are shown in Figure 14. Controlling for performance, a team's win total appears to have no effect on its attendance.

[12] In fact, though none of the performance variables are significant, an F-test shows that the results in Figure 12 present a significantly better fit than those in Figure 11 ($F = 5.25$, $p < 0.001$).

Figure 14. Regression Results, Wins Removed

Variable	β	S.E.	T-Value	P-Value
α	194,636	3,493,965	0.06	0.956
$Wins_{t-1}$	12,634	3,893	3.25	0.001
$Playoffs_{t-1}$	160,409	92,965	1.73	0.086
adj1B	413	3,870	0.11	0.915
adj2B	-9,550	6,081	-1.57	0.117
adj3B	-5,991	9,247	-0.65	0.518
adjHR	-2,905	10,788	-0.27	0.788
adjSB	-1,717	1,956	-0.88	0.381
adjCS	2,543	4,526	0.56	0.575
adjBB	-1,435	2,719	-0.53	0.598
adjHBP	-4,148	3,683	-1.13	0.261
adjField	2,142	897	2.39	0.018
adjBsRun	-3,952	5,048	-0.78	0.434
BaseRuns	7,170	7,593	0.94	0.346
BaseRunsA	-583	460	-1.27	0.206
Expansion	900,397	512,082	1.76	0.080
$Expansion_{t-1}$	146,763	289,339	0.51	0.612
NewStadium	493,654	166,325	2.97	0.003
MarketSize	3,259	542	6.01	0.000

$R^2 = 0.566$
$F = 20.35$
$S.E. = 493{,}382$

Is there a tradeoff between salary and contract length?

Theory suggests that longer contracts should, all else being held equal, result in smaller salaries, since players presumably have a lower risk tolerance than owners, whose risk is reduced by the fact that they employ a roster of 25 players. This situation is analogous to portfolio theory: If player-specific risk is σ, then an owner's risk is equal to $\frac{\sqrt{n \times \sigma^2}}{n} < \sigma$ for all $n > 1$. Empirical results, however, have not always reflected theory.

Meltzer (2005) finds a positive association between salary and contract length even when controlling for performance, except for young and chronically injured players. Krautmann and Oppenheimer (2002) also find a positive association between salary and contract length, though their research does suggest that the returns to performance decrease as contract length increases.

I do not include contract length in the regression reported in Figure 6, as salary and contract length are jointly determined and invariably correlated: Better players tend to receive both longer contracts and higher salaries. The residuals of that regression, therefore, are bound to be negatively correlated with contract length: Players that underperform their salary will also have longer contracts than their performance would indicate, and vice-versa. It is possible to avoid that problem by instead grouping actual and expected salaries by contract length—if there is no tradeoff between salary and contract length, then the two numbers would be expected to match at each length of contract; if, as portfolio theory indicates, owners are more risk-tolerant than players, then players with longer contracts should receive lower salaries than expected. What actually occurs? I report the actual salaries received by players in my dataset, as well as their expected pay, based on the regression in Figure 6 and with a further adjustment for the tendency of that equation to under-predict salary, in Figure 15[13]:

[13] The regression equation under-predicts salaries due to the forced-zero intercept, which is a necessary annoyance. Allowing an intercept in the regression flattens the slope of the relationship between salary and performance, causing the coefficients to be significantly underestimated.

Figure 15. Actual vs. Predicted Salary, Grouped by Contract Length

Years	Total Adjusted Salary	Total Predicted Salary	Actual/Predicted
2	$681,915,843	$515,972,420	32%
3	$623,464,872	$692,526,018	-10%
4	$471,605,999	$508,736,317	-7%
5	$402,331,452	$429,958,908	-6%
6	$111,105,776	$115,675,641	-4%
7	$102,942,831	$108,474,338	-5%
8	$87,667,560	$85,675,711	2%
10	$132,924,946	$156,939,926	-15%

At most contract lengths, actual and predicted salaries match up fairly well, with one glaring exception: Players that received two year contracts out-performed their expected salaries by 32%. That result is significant at $p < 0.001$. Just as in Meltzer (2005) and Krautmann and Oppenheimer (2002), short contracts appear to be associated with lower than expected salaries. As for why, it is unclear. One possibility is that not all players carry the same risks, and riskier players are awarded shorter contracts. Meltzer's finding that chronically injured players appear to receive shorter deals when controlling for performance seems to suggest that conclusion. An interesting direction for future research might be to explore whether players whose performance is inherently subject to more volatility (whether that be based on the volatility of their past performance, or some sort of group-specific characteristic that is associated with additional volatility, such as old age) receive shorter-than-expected contracts. The results presented here cannot speak to such details, but they do strongly suggest that players receiving two year contracts are significantly underpaid relative to their peers.

Is the cost of a win constant for all teams?

One assumption made in this paper is that when a team signs a player, it signs him for the marginal wins he can provide to the club, and that those wins therefore can be treated as a commodity; that is to say, that wins are purchased at a singular market price. This is not to say that all teams will target the same number of wins, but rather that the value of a *marginal* win will be the same for all teams—the number of wins at which marginal value equals marginal cost will vary based on the team's revenue function and the utility its owner derives from winning. (See the theoretical model earlier in this paper for more detail.) That is not necessarily an intuitive assumption: To many, it seems obvious that larger market teams such as the New York Yankees or Boston Red Sox will pay more per win than will teams from smaller markets, especially since they are likely to target the limited number of star players that can provide a significant number of marginal wins. If this were the case, then the theoretical standing upon which the empirical work in this paper is based would be invalid. However, Brown and Jepsen (2009) find no evidence that higher-revenue teams pay more for players than do their lower-revenue counterparts. This is also a question that can easily be re-tested using the results already obtained in this paper.

If large-market teams pay more per win than do small market teams, then we should be able to observe a correlation between the residuals obtained in the regression detailed in Figure 6 and the market size of the signing team, as per Silver (2007). Instead, there is almost no correlation between the two figures ($r = 0.04$, $p = 0.56$). In other words, larger market teams do not appear to pay players more than would be expected based on their statistical performance, confirming Brown and Jepsen (2009) and providing some validation for a model that envisions the MLB free agent market as a market for wins rather than players.

Conclusion

The purpose of this paper was to explore whether Major League Baseball teams properly various categories of player performance relative to their marginal impact on a team's win. Neoclassical labor theory posits that when a worker's productivity can be observed, as is the case in Major League Baseball, his salary will be equal to his marginal product. If that were the case, then major league teams would have to properly value the various ways in which a player can contribute to their win total. I find that for the years 1999-2009, this is not the case. Though teams do appear to appropriately value *most* statistical categories of performance, there exists strong evidence that in the time period examined, major league teams were significantly undervaluing the contribution that hitters made through their fielding—in fact, they seem not have valued it at all. Even in the presence of strong competition and near-perfect information, the MLB free agent market appears to have been inefficient in the years examined.

In addition to the main question of this paper, I examine a number of related questions. In bullet-point form, I find that:

- Contrary to Hakes and Sauer (2006), teams do not appear to have undervalued walks prior to the publication of *Moneyball*, nor did they change how they value walks post-publication. It does appear that teams placed more emphasis on home runs while paying less for singles after the release of *Moneyball*.
- Teams appear to properly weight a player's previous seasons when offering free agent contracts.

- Player performance appears to drive attendance in the same manner as it relates to winning; it does not appear that fans will show up in larger numbers for teams that achieve their wins in specific ways (i.e. by hitting home runs rather than through walking).
- Players that sign two-year contracts appear to be under-paid relative to their multi-year brethren.
- Large-market teams pay the same price per marginal win on the free agent market as do small-market teams, confirming Brown and Jepsen (2009).

These related questions shed further light on the MLB free agent market, while confirming the underpinnings of the theoretical model that I present in this paper.

Though I examine a number of related questions, there are still many avenues for research in this area. Future researchers might want to examine how expected the uncertainty around a player's expected performance affects his salary, model attendance on a game-by-game basis as opposed to the seasonal model examined here, calculate team-specific revenue functions and compare expected marginal product to the actual salaries teams hand out to free agents, or repeat some or all of the tests presented in this paper for other sports or even other professions where such detailed data is available. Though I have tried to be thorough in this paper, there are still many questions on this subject to be examined. Still, the findings in this paper may prove useful to economists as well as major league teams and player agents, if the inefficiencies in the MLB labor market observed here still exist.

References

Benjamin, Amalie. “Defensive-minded.” *Boston Globe* 4 Apr. 2010, Sunday ed., sec. V: 2-3. Print.

Bollinger, Christopher R., and Julie L. Hotchkiss. “The Upside Potential of Hiring Risky Workers: Evidence from the Baseball Industry.” *Journal of Labor Economics* 21.4 (2003): 923-44. Print.

Bradbury, John C. “Does the Baseball Labor Market Properly Value Pitchers?” *Journal of Sports Economics* 8.6 (2007): 616-32. Print.

Brown, Kenneth H., and Lisa K. Jepsen. “The Impact of Team Revenues on MLB Salaries.” *Journal of Sports Economics* 10.2 (2009): 192-203. Print.

Burger, John D., and J.K. Walters. “Uncertain Prospects: Rates of Return in the Baseball Draft.” *Journal of Sports Economics* (2009): 1-17. Print.

Cartwright, Brian. “Oliver, Smarter Than Your Average Monkey.” *The Hardball Times*. 30 Dec. 2010. Web. 14 Apr. 2011. <http://www.hardballtimes.com/main/article/oliver-smarter-than-your-average-monkey/>.

Clayton, Matthew, and David Yermack. “Major League Baseball Player Contracts: An Investigation of the Empirical Properties of Real Options.” 1999. MS 99-051. New York University, New York.

Craggs, Tommy. “MLB Confidential: The Financial Documents Baseball Doesn’t Want You To See, Part 1.” *Deadspin*. 23 Aug. 2010. Web. 07 Mar. 2011. <http://deadspin.com/#!5615096>.

Einbinder, Benjamin. “An Analysis of Final-Offer Arbitration Outcomes for Batters in Major League Baseball from 2002-2006.” Thesis. Haverford College, 2007. Print.

Fields, Brian. “Estimating the Value of Major League Baseball Players.” Thesis. East Carolina University, 2001. Print.

Fort, Rodney. “The Value of Major League Baseball Ownership.” *International Journal of Sports Finance* 1 (2006): 9-20. Print.

Gassko, David. “Do Chicks Dig the Long Ball?” *The Hardball Times*. 31 Jan. 2008. Web. 10 Mar. 2011. <http://www.hardballtimes.com/main/article/do-chicks-dig-the-longball/>.

Gregory, Sean. “Why Has 2010 Been the Year of the Pitcher in Baseball? - TIME.” *Time*. 31 July 2010. Web. 07 Mar. 2011. <http://www.time.com/time/nation/article/0,8599,2007984-1,00.html>.

Hakes, Jahn, and Chad Turner. “Pay, Productivity, and Aging in Major League Baseball.” *Journal of Productivity Analysis* 35.1 (2011): 61-74. Print.

Hakes, Jahn K., and Raymond D. Sauer. “An Economic Evaluation of the Moneyball Hypothesis.” *Journal of Economic Perspectives* 20.3 (2006): 173-85. Print.

Hakes, Jahn K., and Raymond D. Sauer. “The Moneyball Anomaly and Payroll Efficiency: A Further Investigation.” *International Journal of Sports Finance* 2.4 (2007): 177-89. Print.

Heipp, Brandon. “Base Runs.” *Buckeyes and Sabermetrics*. Web. 10 Mar. 2011. <http://gosu02.tripod.com/id108.html>.

Kahn, Lawrence M. “Free Agency, Long-Term Contracts and Compensation in Major League Baseball: Estimates from Panel Data.” *The Review of Economics and Statistics* 75.1 (1993): 157-64. Print.

Krautmann, Anthony C., and Margaret Oppenheimer. “Contract Length and the Return to Performance in Major League Baseball.” *Journal of Sports Economics* 3.1 (2002): 6-17. Print.

MacDonald, Don N., and Morgan O. Reynolds. “Are Baseball Players Paid Their Marginal Products?” *Managerial and Decision Economics* 15.5 (1994): 443-57. Print.

Meltzer, Josh. “Average Salary and Contract Length in Major League Baseball: When Do They Diverge?” Thesis. Stanford University, 2005. Print.

Miceli, Nicholas S., and Alan D. Huber. “‘If the Team Doesn't Win, Nobody Wins:’ A Team-Level Analysis of Pay and Performance Relationships in Major League Baseball.” *Journal of Quantitative Analysis in Sports* 5.2 (2009). Print.

Perry, Dayn. “Do Players Perform Better in Contract Years?” *Baseball Between the Numbers: Why Everything You Know About the Game Is Wrong*. New York: Basic, 2006. 199-206. Print.

Ruane, Tom. “The Value Added Approach to Evaluating Performance.” *Retrosheet*. 4 July 2005. Web. 07 Mar. 2011. <http://www.retrosheet.org/Research/RuaneT/valueadd_art.htm>.

Scully, Gerald W. “Pay and Performance in Major League Baseball.” *The American Economic Review* 64.6 (1974): 915-30. Print.

Silver, Nate. “Lies, Damned Lies: Defining a Market, Part Two.” *Baseball Prospectus*. 4 May 2007. Web. 07 Mar. 2011. <http://www.baseballprospectus.com/article.php?articleid=6184>.

Silver, Nate. “Lies, Damned Lies: Rethinking Replacement Level.” *Baseball Prospectus*. 22 Mar. 2006. Web. 07 Mar. 2011. <http://www.baseballprospectus.com/article.php?articleid=4891>.

Smith, Sean. *Baseball Projection*. 19 Jan. 2010. Web. 07 Mar. 2011. <http://baseballprojection.com/>.

Swartz, Matt. “Ahead in the Count: The Cost of OPP.” *Baseball Prospectus*. 17 May 2010. Web. 07 Mar. 2011. <http://www.baseballprospectus.com/article.php?articleid=10883>.

Tarman, Andrew. “Does the Arbitration Process Solve Monopsonistic Behavior in Baseball?” *The Park Place Economist* 13 (2005): 21-28. Print.