Regime Switching and Wages in Major League Baseball
by
Michael J. Haupert and James Murray
Department of Economics, University of Wisconsin – La Crosse

For presentation at the WEAI meetings July 2010
June 2010 Draft. Please do not cite without permission of authors.

Abstract

Over the course of the 20th century American wages increased by a factor of about 100, while the wages of professional baseball players increased by a factor of 450. But that increase was neither smooth nor consistent. In this research we examine the causes of the wage growth with a regime switching regression model applied to a unique data set of baseball salaries. Regime switching is one of several methods in time series to allow for time-varying coefficients. The method allows individual regression coefficients to take on one of two or more values (or regimes) as time progresses, and there can be multiple instances of switching from one value to another throughout the sample period. We focus on a two regime case, so coefficients can switch between high and low values. We impose further structure to our application to Major League Baseball, such that when a regime change occurs, it applies to all players at the same time. We use this model to predict the growth rate of a player's salary based on past experience and productivity, among other controls. A switch detected in the regression coefficient on productivity, occurring to all players in the league symmetrically, may imply a change in rules, a change in the league's competitiveness, or a change in the relative bargaining power between players and teams. We assume factors that may cause such a regime switch are unknown or unobservable, so it is the dynamics of the dependent and explanatory variables that determine the number of regime switches, and the relative difference in coefficients between one regime and the other. We compare the predictions for regime switches to recognized dates of significant change in market conditions and institutional structure. We also compare the pattern of change in baseball salaries to general wage growth in order to explore global vs. industry specific patterns of regime switching.

I. Introduction

Professional baseball players are among the highest paid workers in the American economy. Their minimum salaries are several times that of the average American salary, and the average wages are an even greater multiple (Figure 1). They enjoy a minimum salary of $400,000 per year, an average salary exceeding $2 million per year, guaranteed contracts, and arguably the strongest union in American history. Stories of the financial escapades of professional baseball teams and the salaries they pay their employees are common in the sports and business press.

It was not always this way however. It was not until the mid 1970s when professional baseball players blazed the trail for all professional athletes by gaining the right to bargain competitively for their wages. Before that time the players were bound to their original employer for the duration of their careers. The labor market for ballplayers was a classic monopsony.

There have been numerous studies of the baseball player labor market. These previous works have focused on issues such as labor exploitation, the effect of competitive labor markets on competitiveness in the league, and the impact of competition for players on league structure and financial stability. This earlier research has produced some interesting results, including models to calculate the financial contribution of individual players to their teams and the degree of labor exploitation. In this paper we use a regime switching regression model to investigate the changing relationship between salary and performance over time.

So what determines the wage of a player? In this article we confine our study to pitchers. How did an owner decide what to pay a player in the early years, given that there were no minimum wages and no market to determine the players salary? Evidence suggests that wages were not simply raised on an annual basis for all players at steady rates.[1] It is possible that certain performance characteristics, beyond wins and losses, appeal to fans. We also add experience, measured as number of years in the league, and experience squared, to capture the effect of aging on deterioration of skills.[2]

When competition doesn't exist, what determines wages? Players were not all paid the same salary, yet market pressures surely did not force owners to pay players salaries

[1] See Haupert (2009) for a discussion of MLB wages during different labor regimes.
[2] MacDonald and Reynolds (1994) find both experience variables highly significant in their work.

commensurate with their marginal revenue product. Even though players were not paid their marginal revenue product, it seems reasonable to assume that owners were cognizant of their value when paying their wages. To that end the literature has focused on discovering the relationship between wages and marginal revenue product to get at the root of the question, what determines wages? The variables that affect wages can be categorized in three basic areas: player performance variables, macroeconomic variables, personal characteristics of players. Ultimately, the value of a player, which will be reflected in his wage, is related to his impact on team revenue.

Historical Background

From its earliest days, Major League Baseball has been a monopsonistic employer. Beginning in 1876 with the founding of the National League of Professional Base Ball Clubs (NL) the industry was designed such that employers (team owners) minimized the ability of employees (players) to sell their services competitively. The name chosen for this new league was significant because up until this time, all baseball organizations had been player associations. Now, as the name implied, players were to be employees of a club and members of a league.

The new league "provided for a new order, ingeniously designed to be nourished on both monopoly and competition. Under its aegis, member clubs were to compete with each other for renown and receipts, but only within the confines of a prescribed pattern." [3] Ultimately both stability and the disruption of bitter labor disputes would result from the formation of the National League. Stability would bring about a superior brand of baseball, considered the best in the world. Labor problems were key in the formation of competing leagues.[4] All failed, though the American League ultimately merged with the National League to form what is to today known as Major League Baseball.

For much of the first century after the formation of the National League the players had little say in any affairs of their clubs and no representation in the governing institutions of the League. Club executives simply presented contracts to their players, refusing to negotiate. Different clubs used various measures to keep players in line, including sobriety regulations,

[3] Seymour, p. 85.

[4] Failed competitors and their years of operation: Union Association 1884, American Association 1882-91, Players League 1890, Federal League 1914-15. The American League was formed as a competing league in 1901 and merged with the National League in 1903.

medical examinations by club doctors and suspensions for poor play, illness or insubordination. If a player violated any club directives, he could even be "blacklisted" from professional baseball until he repented.

Team owners formalized their control over the players with the adoption of the reserve clause following the 1879 season. Owners agreed that each of them would "reserve," or keep off the market, five players of their eleven man rosters for the following year. The list of reserved players was circulated among all owners, and each agreed not to employ or even negotiate with any other team's reserved players. The plan was initially designed to prevent rich teams from acquiring all the best players. The number of reserved players slowly increased over the years until entire club rosters were designated as reserved. Beginning in 1887 the list became a clause in the standard player contract. The reserve clause remained intact for nearly a century, turning the baseball labor market into a tightly controlled monopsony.

With baseball players bound to one employer in perpetuity by the reserve clause, the labor market for ballplayers was a classic monopsony. Players were free to negotiate with prospective employers only for their first contract, usually as a young man, in many cases even a minor. At this age the prospects for any player were extremely uncertain. In addition, most unschooled young potential ballplayers lacked both the sophistication and the knowledge of the market to do much, if any, bargaining. As a result, the first contract usually imparted little to the ballplayer in the way of bargaining leverage.

Once signed to a professional contract, players lost control of their fate as a professional baseball player unless they were released from that contract at the discretion of the team. The reserve clause gave teams the ability to renew a player's contract in perpetuity on terms dictated by the team. The sale of contracts from one team to another was a common method of raising revenue for the sellers and building a better club for the buyers. Consistent in all of this was the lack of any input by the player. His only choice was to play for the new team or look for a job in a different line of work.

In this environment the only leverage a player could exert was to threaten to hold out his services. For most players this was not a realistic advantage, as the monopoly MLB held down the number of viable franchises in the league, and hence the number of employed ballplayers. As a result, there were many players in lesser leagues scattered around the country who were viable substitutes for the player who held out his services. A holdout strategy was only possible for the highest quality players at the peak of their career, for whom a viable substitute did not exist. Even then, holdouts were rarely successful, as the opportunity cost of a holdout was much higher for the players than the teams.

Joe DiMaggio, for example, staged a hold-out in the spring of 1938. At the time he was arguably the best baseball player in the world. He was a celebrity and a fan favorite with enormous drawing power. If anybody should have been able to execute a successful holdout, it would have been Joe DiMaggio in 1938. The Yankees were coming off back-to-back World Series championships and their highest attendance in seven years. Not since Ruth retired had the team done so well on or off the field. For his contribution, DiMaggio had just put together two of the finest seasons in baseball history, and the fact that they were his first two seasons made this performance all the more remarkable. Connie Mack, the venerable owner of the Philadelphia Athletics and a fifty year baseball veteran, called him the greatest drawing card in the game. DiMaggio was a full-fledged superstar.

When he asked for a raise from $15,000 to $40,000 it seemed to him like a reasonable request. In fact, his marginal revenue product (MRP) that year was approximately $400,000, so even at $40,000 the club would be getting a steal.[5] However, DiMaggio discovered that management was holding all the cards. Not only was the reserve clause working against him, but so was the press, and as a result, the fans.

DiMaggio threatened to stage an indefinite holdout, but three days after the season started, he caved in. Yankee owner Jacob Ruppert triumphantly shared his telegram with the press: “Your terms accepted,” adding that DiMaggio’s salary would be docked each day until his manager deemed him ready to play. That turned out to be nearly two weeks and almost $1500 into the season. Perhaps worse than the lost salary was the lost adulation of the fans, who had been turned against DiMaggio.

Besides the twin cannons of the reserve clause and the press applying pressure on a holdout, the teams could rely on their monopoly control of the industry to always have a stockpile of available players on hand to bring to the club in place of a malcontent or fading talent. The Yankees threatened that if the great DiMaggio held out, they would simply proceed with backup Myril Hoag in his place. Throughout the spring Hoag batted .352, lending credence to that threat.

---

[5] Haupert (2009)

This type of market structure changed dramatically in 1975 with the elimination of the reserve clause. The growing power of the Major League Baseball Players Association, formed in 1952, led to a collective bargaining agreement in 1968, and the 1970 ruling by the National Labor Relations Board that MLB and the MLBPA use outside arbitrators for resolving grievances. This ultimately led to free agency when an arbitrator ruled in 1975 that the reserve clause was not indefinite, creating the first true free agents in 1976. The players and owners negotiated the present system of restricted free agency which gives owners the exclusive right to bargain with a player for the first six years of his contract. Thereafter a player is free to negotiate with any team. In its original iteration players were allowed to negotiate with only a limited number of clubs, but this quickly gave way to an unfettered market, in which players were granted the right to sell their services to the highest bidder. Players are still bound to their teams for the first six years of their contract before becoming eligible for free agency. However, after two years they are eligible to file for arbitration. The forma of arbitration system MLB and the MLBPA have created is known as final offer arbitration. In the case of an impasse in salary negotiations, each party may file a salary claim with the arbitrator, who after a hearing in which each side presents their case, must choose one or the other salary. Oftentimes the two parties agree to a compromise salary in between the time the claims are filed with the arbitrator and the scheduled hearing. Prior to this eligibility for arbitration, players are subject to the salary dictates of their teams, though there is a minimum salary and a maximum allowable salary reduction from year to year.

What determines wages: literature review

The seminal work in the area of ballplayer wages is really that of Gerald Scully, who first attempted to "crudely measure the economic loss to the players"[6] as a result of the reserve clause. Scully used limited salary data from the 1968 and 1969 seasons to calculate the MRP of players and compare that to actual salaries. He looked at both hitters and pitchers, but our interest is only in pitchers. He regressed the log of salary on lifetime percentage of innings pitched and K/BB plus a variable for experience. Scully notes that he employed a variety of performance measures in numerous regressions, but "the fact that one performance measure or another or one plausible effect or another does not appear in the regression equations reported

[6] Scully (1974), p 915

here does not mean that the measure or the effect was not associated with salary variations." [7] In fact, most of the variables he considered were highly correlated with players' salaries, but their unique effects could not be isolated. For pitchers, these other variables included games won per full season played, innings pitched per full season, career games won, career innings pitched, seasons pitched and differences between performance last year and lifetime performance. The only variable he could not find significant was ERA. In a later study (Scully 1989) he incorporated free agency into his lifetime K/BB and lifetime IP% variables. He found lifetime K/BB ratio to be the single best measure of pitching performance and noted that player salaries tended to rise automatically over time with experience independently of average lifetime performance. In general, he finds four factors that are most important in determining salary: player performance, the weight of the player's contributions to team performance, years of MLB experience, and the enhanced bargaining power of being a superstar.

Scholars following Scully tweaked his MRP or team revenue equations, developed additional or alternate measures of player performance that they regressed on player salaries, or asked slightly different questions based on the concept of measuring the relationship between player performance and salary.

In a more general study of whether workers are paid their MRP, Robert Frank (1984) finds that within firms, wage rates vary substantially less than do individual productivity values.[8] Frank claimed that many firms, because of the difficulty in monitoring, tended to follow wage formulas linked to experience, education and firm tenure despite differences in worker productivity. He claims that employees may prefer this type of compensation as a means of smoothing income over time rather than the vagaries of annual income changes tied to productivity changes. This is interesting if one assumes that it is easier to give a small annual raise to a player than to increase his salary substantially after a good year and then to decrease it correspondingly after a bad year.

Frank notes that "it is often suggested that 'equity considerations' account for why internal wage structures are so much more egalitarian than the ones predicted by the marginal productivity theory of wages. . . but it raises the question of why, if equity is truly what they seek, do the best workers in existing firms not join new firms with other workers who are just as

[7] Scully (1974) p 934 fn
[8] Frank (1984) p 549

productive as themselves?"[9] The explanation in baseball is quite simple: until 1976, the reserve clause prevented this. And after 1976 the popular opinion is that the richest teams do exactly this: pay the highest salaries to the best players (read New York Yankees).

Zimbalist (1992b) addresses Franks's hypothesis by breaking his sample into experience categories consistent with player bargaining leverage. Players with less than three years experience, who were not eligible for arbitration or free agency, comprise the first group. The second group had 3-5 years of experience. These players qualified for arbitration but not free agency. The final group had six or more years experience, thus qualifying for free agency. He found a strong positive correlation between salary and experience.

Krautmann and Openheimer (2002) and Kahn (1993) and MacDonald and Reynolds (1994) incorporated contract duration and bargaining environment into the equation, regressing salaries on length of contract and arbitration eligible, free agent eligible, and final year of contract.

MacDonald and Reynolds (1994) use runs scored and ERA instead of the slugging average and walks-to-strikeout ratio measures used by Scully. They also incorporated a dummy variable taking on the value of one if the team's winning percentage fell below .500 averaged over the two previous seasons.

Hoaglin and Velleman look at several approaches to the salary data issue. They note that experience is not a linear relationship and needs to be addressed either as a quadratic, or even better as a piecewise linear pattern, where salary grows linearly for the first seven years of experience and levels off after that.

Fort considers the relationship between age and experience and treats it nonlinearly. He finds that age and experience are related, but different, and both affect players greater at early and later stages of their career. In other words, an extra year of age and experience are more significant to younger players (positively) and older players (negatively). He also argues for recent versus career-to-date performance as important variables. IP and K/BB are significant as is the distinction of being a starting pitcher.

Burger and Walters supply evidence for the importance of macro variables in salary determinations. They find that market size and team performance interact to increase the MRP and thus the salary of players in large markets and winning teams. They estimate that teams in

[9] Frank (1984) p 569

the largest markets derive as much as six times more MR from additional win, thus they are willing to pay more for free agents.

Krautmann (1999) hit upon the idea of positional importance and found dummy variables for shortstops and catchers significant in a study of hitters. He carried that concept on in a 2003 paper with Gustafson and Hadley when he divided the pitching sample into starters, long relievers and closers. They conclude that aggregating pitchers leads to false outcomes. When aggregating pitchers they found team total revenue, lefty dummy, ERA, W, SV, IP, K/IP and dummies for relievers and stoppers all significant. When disaggretating the data into three pitching subspecialties the results differ. For starting pitchers ERA and K/IP prove significant. For long relievers it is wins and K/IP plus lefty dummy and team TR. For closers it is ERA, SV and IP.

The consistency in these studies is their agreement that experience and bargaining status matter in wages and that performance variables should be lagged. There is some variation on which performance variables to use, and each study finds different significant experience variables, which is not surprising given Scully's initial analysis. More recent work has focused on disaggregating player samples by subspecialty. One other similarity in the studies is the very short time periods each looked at. Most have used time periods ranging from one to five years. For used a ten year period. Haupert (2009), using data from the New York Yankees financial records has looked at a longer time period and a novel performance variable, win shares.

While based on the findings of the previous research, this work seeks to both expand and differentiate from the previous literature. First of all, we are looking at a much longer time series with many more observations, crossing many potential regime changing periods. Our dataset begins in 1911 and goes through 2008. Our model is looking to identify not so much what caused salaries to fluctuate, but when the significant variables in determining wages changed.

Salary Data

We use data on salary, age, experience, and performance statistics for 1,300 pitchers from 1911 through 2008, for a total of 4,745 observations. Figure 4 shows plots of the average salary over the sample period, broken into two sub-samples: 1911-1972, which we denote as the ``Reserve Clause Era'', and 1972-2008, which we denote as the ``Free Agency Era''. Even though free agency started in 1976, there is a distinct permanent increase in the growth rates of pitcher occurring in this year. Prior to 1972 the average annual growth rate of nominal salaries was

3.2%. After 1972 the average annual growth rate was 11.7%. Baseball historians will recognize 1972 as the year of the first players strike. While it lasted only two days, it was a watershed event in the balance of power between players and owners.

The top panels in Figure 4 show plots of the actual nominal salary, and the second row of figures show the same data, but with a log-scale. Consistent OLS and maximum likelihood estimators requires stationary data, so we transform the salary data by removing the trend growth rate. Let $y_{i,t}^{obs}$ denote the observed salary for player i and time t (ranges from 1911 to 2008), and let $y_{i,t}$ denote the transformed detrended salary. The detrended salary we use for estimation is given by,

$$y_{i,t} = \frac{y_{i,t}^{obs}}{(1+g)^{(t-1905)}}, \tag{1}$$

where $g = 0.032$ is the growth rate prior to 1972 and $g = 0.117$ is the growth rate from 1972 onwards. The bottom graphs in Figure 4 show plots of this detrended data. It is clear from these plots there have been large movements in baseball pitchers' salaries away from the trend growth line. The purpose of this paper is to determine how and whether we can account for these large movements.

Regime Switching Regression

We employ a regime switching regression model to investigate how the relationship between performance and salary has evolved over time. A regime switching regression model allows more than one regression equation to describe the data, where in a given time period, one regression line may be describing the data, but in the next time period, there may be an exogenous switch to another regression line. This method is appropriate when the relationship between dependent and explanatory variables varies over time for reasons that are either unmeasurable or unknown. Consider the following standard (no switching) pooled panel regression model,

$$y_{i,t} = x_{i,t}'\beta + e_{i,t}, \quad e_{i,t} \sim \mathrm{N}(0,\sigma^2), \tag{2}$$

where subscript i denotes a given individual, subscript t denotes a given time period, $y_{i,t}$ is the observed value for the dependent variable for individual i at time t, $x_{i,t}$ is a vector of explanatory variables that can include a constant for individual i at time t, β is a vector coefficients, and $e_{i,t}$ is an independently normally distributed error term with variance given by σ^2.[10] For our application, $y_{i,t}$ denotes (detrended) salary; and $x_{i,t}$ includes a constant, a player's age, a player's years of experience, and the previous year's performance variables. The performance variables for pitchers are earned runs average (ERA), wins as a percentage of games in the season ($Wins$), saves as a percentage of games in the season ($Saves$), the ratio of strike-outs to base on balls (SO/BB), and the sum of wins and hits divided by innings pitched ($WHIP$). The coefficients in β is then the relative importance of each one of these variables for determining a player's salary.

A single regression line, that is a single set of estimates the coefficients in β and the variance σ^2, can be estimated using ordinary least squares (OLS). We extend this model by considering the following regime switching panel model,

$$y_{i,t} = x_{i,t}'\beta(s_t) + e_{i,t}, \quad e_{i,t} \sim \mathrm{N}(0, \sigma^2(s_t)), \tag{3}$$

where $s_t \in \{1,2,.., S\}$ denotes which state, or regime, the regression relationship is in at time t, and S is the possible number of regimes. The regression coefficients are given by, $\beta(s_t) = \beta_k$ if and only if $s_t = k$; and the variance is given by $\sigma^2(s_t) = \sigma_k^2$ if and only if $s_t = k$. To keep the estimation procedure tractable, we consider a model with only two regimes ($S = 2$). When the league is in regime 1, the regression relationship is characterized by parameters β_1 and σ_1^2; when the league is in regime 2, the regression relationship is characterized by parameters β_2 and σ_2^2.

We assume the probability of being in each regime in each time period evolves according to a Markov process. That is, it evolves exogenously, and depends only on which regime the league was in during the previous time period. Let the probability the league is in regime 1, given it was in regime 1 in the previous period, be given by $p \in (0,1)$; and the probability the league is in regime 2, given it was in regime 2 in the previous period, be given by $q \in (0,1)$. These two Markov probabilities need to be estimated along with the regression coefficients and error term variance for each regime.

[10]For the standard pooled panel regression model, the assumption that the error term is normally distributed is not necessary. We make this assumption at the introduction of the model because it will be necessary in order to estimate the regime switching panel model by maximum likelihood.

Note, s_t has no i subscript, which implies all players in the sample are in the same regime during a given year. Therefore changes in regime, i.e. changes in the regression coefficients determining players salaries, change for all players in the same way at the same time. Such changes might happen for a number of reasons: changes in the league's competitiveness can change how players are rewarded, downswings and upswings in the economy can impact players salaries, or changes in players bargaining power can influence the importance of performance variables. Finally, it is possible there have been changes in the way players are compensated, but the literature may not yet have answers, or many not yet have even detected them.

We estimate the pooled panel regression model using both OLS and regime switching. Hamilon (1989) describe a maximum likelihood procedure for estimating a time series regime switching model, which we extend to a pooled regression panel as described in the appendix.

Results

Table 1 shows the estimates for the regression coefficients under OLS and Regime Switching for each sub-sample. Detrended salaries for the Reserve Clause Era are expressed in thousands of dollars and the detrended salaries for the Free Agency Era are expressed in tens of thousands of dollars. The OLS model is a single equation regression model, so there is no regime switching. The estimates in the columns labeled ``Regime One'' are the coefficients under Regime 1, but the estimates in the columns labeled ``Regime Two (Change)'' are the differences in the regression coefficients from regime 1 to regime 2. For example, the coefficient for the reserve clause era in Regime 1 for wins is 20.5015 and the change occurring during Regime 2 is 15.3359. This implies for each additional win in Regime 1, players are paid on average an additional $20.5015 \times \$1{,}000 = \$20{,}501.50$. To find how each additional win affects pay in Regime 2, add the estimates in each column to find $(20.5015 + 15.3359) \times \$1{,}000 = \$35{,}837.40$. We express the estimates in this way so that statistically significant values in the Regime Two columns indicate a statistically significant change in this coefficient.

As we look down the OLS column and Regime One column during the Reserve Clause Era, we see that age, experience, wins, SO/BB, and WHIP, all significantly explain pitchers' salaries. Looking down the Regime Two column, we see there were significant changes in the coefficients on two variables: experience and wins. The direction of the impact on salary from both variables is the same. When the league was in Regime 2, players were more highly rewarded for their years of experience and their past year's winning percentage, both of which may be signals of relatively high performance over the next year.

The regime switching and OLS regressions are run separately for the Free Agency sub-sample. The OLS results show that every single explanatory variable is statistically significant and is the expected sign. For the regime switching regression, what characterizes a regime change in the Free Agency Era is not necessarily the same things that characterized a regime change in the Reserve Clause Era. Looking down the Regime Two column, we see there are statistically significant changes in every coefficient from Regime 1 to Regime 2. In every case but one – saves, the regression coefficients become more extreme. The coefficients on age and ERA are negative as expected in Regime 1. In Regime 2, they decrease even further, so there is a greater penalty on salary for ERA and age. All other coefficients are positive in Regime 1 and become even larger in Regime 2. Signals of positive player performance are more highly rewarded in Regime 2. Finally, the constant term is also higher in Regime 2, which indicates players are also being paid more on average, regardless of age, experience, and past performance. The one exception previously mentioned is saves. In Regime 1 players are rewarded for more saves, but the regression coefficient drops close to zero in Regime 2 ($18.8307 - 20.8688 = -2.0381$). With the given the standard errors, this is not statistically significantly different from zero.

Figure 5 shows the estimated probability the league was in each regime for each year in the sample.[11] The plots show the probability the league is in Regime 2, so if it is close to zero there is strong evidence the league is in Regime 1 and if it is close to one, the league is likely in Regime 2. The years to the right of the graphs report when the probability exceeded 0.5. Using this information on the evolution of the regimes and the estimation results in Table 3, we now need to address this switching behavior that cannot be captured with our explanatory variables and standard constant coefficient regression models.

When considering the reserve clause era, regime switching occurs at some rather obvious major economic breakpoints, such as the Great Depression and the conclusion of WWII. It is the change in the free agency period which is of greatest interest at this point. We observe a long lasting shift from regime 1 to regime 2 from 1991 through 2004, highlighted by the large, significant shift in the importance of saves in predicting pitcher salaries. This suggests a change

[11]These are estimated using the smoothing method described in Hamilton (1989).

in the way that baseball teams regarded relief pitching, or at least the way they compensated it. While we have no definitive answers to this change, we do have some theories which will guide our future research in this regard.

The period from 1988-1993 is when the major changes occurred in the deployment of relief pitchers. That's when the move to the one-inning save took place, which might have impacted salaries because of the perception that "closers" did not have to work as hard for their saves. More attention was being paid to the other short relievers, and this was a time when they were starting to get agents of their own. Before that, the guys who pitched in the 6th and 7th innings were for the most part anonymous. This is the time during which the “hold” was introduced. It has never been defined to the satisfaction of baseball statisticians, and is not an official statistic. But, it arrived in conjunction with agents who had a vested interest in creating a measure that would make their new clients’ contributions look like something tangible and measurable.[12] Overall, these changes in the perception of the value of different types of pitchers may have impacted the role of saves in salary determination.

Offences exploded during the period 1991-2004, most notably after the 1993 and 1997 expansions. With heavy offense, pitching statistics in general declined, and it might be that in salary negotiations and/or arbitration, it was easier to make a case that despite the number of saves, the player just didn't pitch that well.

Off the field changes were occurring during this period as well. In 1990 a 32 day lockout at the beginning of spring training ended with a new collective bargaining agreement between the players and the owners. One feature of the new agreement was a shortening of the time a player had to wait until he was eligible for arbitration. That same year an independent arbitrator ruled in favor of the players in the fourth collusion case (the players had also won the previous three cases). The players were awarded $102 million dollars and several were freed from their existing contracts at the conclusion of the season when the arbitrator ruled that owners had colluded to repress player salaries during the 1987-88 off season. These financial setbacks to the owners may also have contributed to the way in which they evaluated player salaries.

Conclusion

We employ a substantially larger and longer running data series and use a well tested set of performance and experience variables to look at baseball player salaries. Employing a regime

[12] Schechter (2006)

switching model, we identify several periods during the 20^{th} century when compensation relationships change. Some of these changes seem rather obvious, such as the Great Depression and WWII. The regime change from 1991-2004 gives us greater pause for consideration. We have hypothesized several potential explanations for this, and now must embark on a further research agenda to test them.

Appendix A

Consider the following pooled regression model with regime switching,

$$y_{i,t} = x_{i,t}'\beta(s_t) + e_{i,t}, \tag{1}$$

where subscript i denotes a given individual, subscript t denotes a given time period, $x_{i,t}$ is a vector of explanatory variables that may include both variables that vary across time for an individual and variables that remain constant over time. The regime state is given by $s_t \in \{1,.., S\}$ where S is the number of regimes. The vector of coefficients is given by $\beta(s_t) = \beta_k \; if \; s_t = k$, and the error term is independently and identically normally distributed, $e_{i,t} : N[0, \sigma(s_t)]$, where the standard deviation is given by $\sigma(s_t) = \sigma_k \; if \; s_t = k$. The regime state, s_t, evolves according to the Markov chain, $P(s_t = k \mid s_{t-1} = j, \Psi_{t-1}) = p_{jk}$, where p_{jk} denotes the probability the economy switches from state j to state k as time enters period t and is another parameter to be estimated along with the other regression parameters, and Ψ_{t-1} simply denotes all information up through period $t-1$.

Given the error term $e_{i,t}$ is normally distributed, if $s_t = k$ was known, the probability density function for $y_{i,t}$ is given by,

$$f(y_{i,t} \mid s_t = k, \Psi_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left\{-\frac{(y_{i,t} - x_{i,t}'\beta_k)^2}{2\sigma_k^2}\right\}. \tag{2}$$

The filtering method for evaluating the unconditional joint density function, $f(y_t \mid \Psi_{t-1})$ mirrors quite closely the iterative method suggested by Hamilon (1989) , where y_t denotes the set of observations for every individual at period t, $y_t \equiv \{y_{1,t}, y_{2,t},..., y_{n_t,t}\}$, and n_t is the number of individual for which data is available at time t. Each iteration begin with the input $P(S_{t-1} = j \mid \Psi_{t-1})$ for every $j \in \{1,..., S\}$ and has the output $P(S_t = k \mid \Psi_t)$ and the process requires an initial condition for $P(S_0 = j)$. The filtering procedure takes as given the parameters β_k, σ_k, and p_{jk} for all j,k. Maximum likelihood estimates for these parameters can be obtained by maximizing the the joint density function for all the data (the output from the filtering procedure) with respect to these parameters. The filtering algorithm follows these steps:

• Step 1: Find probabilities for being in each regime in time t, given information up through period $t-1$. These probabilities are given by,

$$P(s_t = k \mid \Psi_{t-1}) = \sum_{j=0}^{S} P(s_t = k \mid s_{t-1} = j) P(s_{t-1} = j \mid \Psi_{t-1}),$$

where $P(s_t = k \mid s_{t-1} = j) \equiv p_{jk}$ is the Markov switching parameter, and $P(s_{t-1} = j \mid \Psi_{t-1})$ is known from the previous iteration (or initial condition).

• Step 2: Evaluate the conditional joint density function $f(y_t \mid \Psi_{t-1})$ which is computed by evaluating the following successive densities:

$$f(y_t \mid s_t = k, \Psi_{t-1}) = \prod_{i=1}^{n_t} f(y_{i,t} \mid s_t = k, \Psi_{t-1}),$$
$$f(y_t \mid \Psi_{t-1}) = \sum_{k=1}^{S} f(y_t \mid s_t = k, \Psi_{t-1}) P(s_t = k \mid \Psi_{t-1}).$$

The first equation is valid since $e_{i,t}$ and $e_{i',t}$ are independent for $i \neq i'$, and the density $f(y_{i,t} \mid s_t = k, \Psi_{t-1})$ is given in equation (2). In the second equation $P(s_t = k \mid \Psi_{t-1})$ is given from step 1.

• Step 3: Evaluate the updated probability for being in each regime in time t, given information up through period $t-1$. These probabilities are given by,

$$\begin{aligned} P(s_t = k \mid \Psi_t) &= P(s_t = k \mid y_t, \Psi_{t-1}) = \frac{f(y_t, s_t = k \mid \Psi_{t-1})}{f(y_t \mid \Psi_{t-1})} \\ &= \frac{f(y_t \mid s_t = k, \Psi_{t-1}) P(s_t = k \mid \Psi_{t-1})}{f(y_t \mid \Psi_{t-1})}, \end{aligned}$$

where the densities and probability needed to evaluate the second line are given in steps 1 and 2.

• Step 4: Return to step 1 until $t = T$, where T is the number of periods in the sample.

The joint distribution for all the data is given by,

$$f(y^T \mid \Psi_{T-1}) = \prod_{t=1}^{T} f(y_t \mid \Psi_{t-1}), \tag{3}$$

where $f(y_t \mid \Psi_{t-1})$ is given from step 2. Taking logs, this can be transformed to the log-likelihood function,

$$l(y^T) = \sum_{t=1}^{T} \log \left(f(y_t \mid \Psi_{t-1}) \right) \tag{4}$$

Numerical maximization methods can be used to maximize equation (4) to obtain maximum likelihood estimates for $\beta(s_t)$ and $\sigma^2(s_t)$ and unique Markov probabilities $p_{j,k}$.

Figure 1

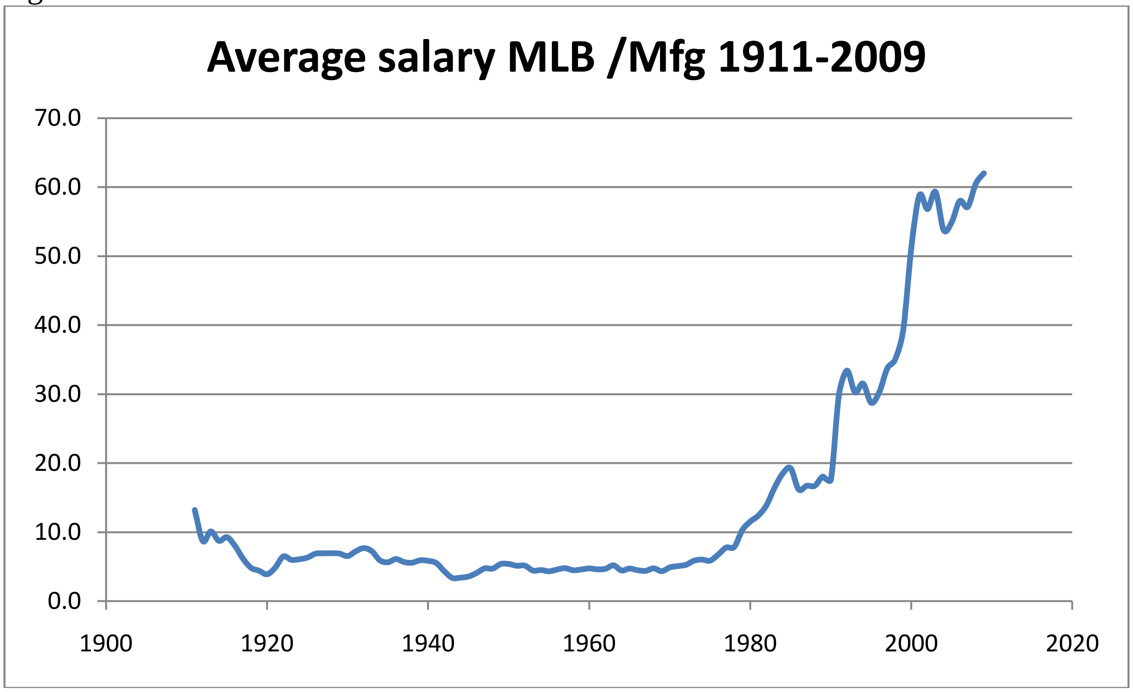


Figure 2

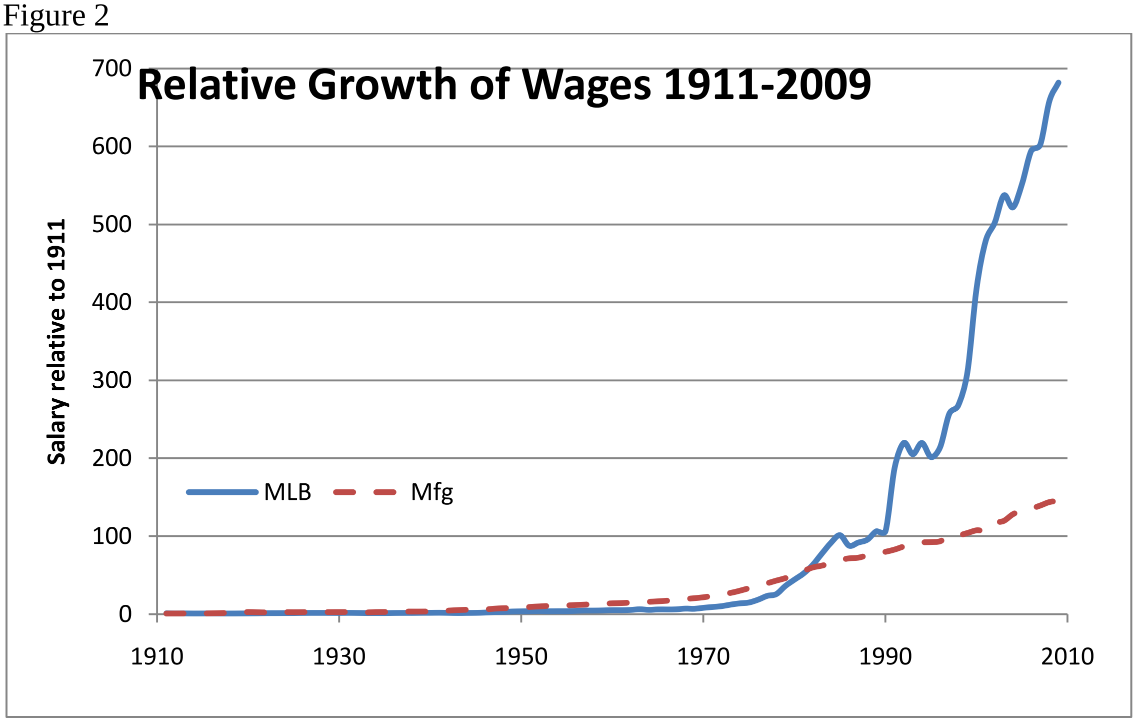
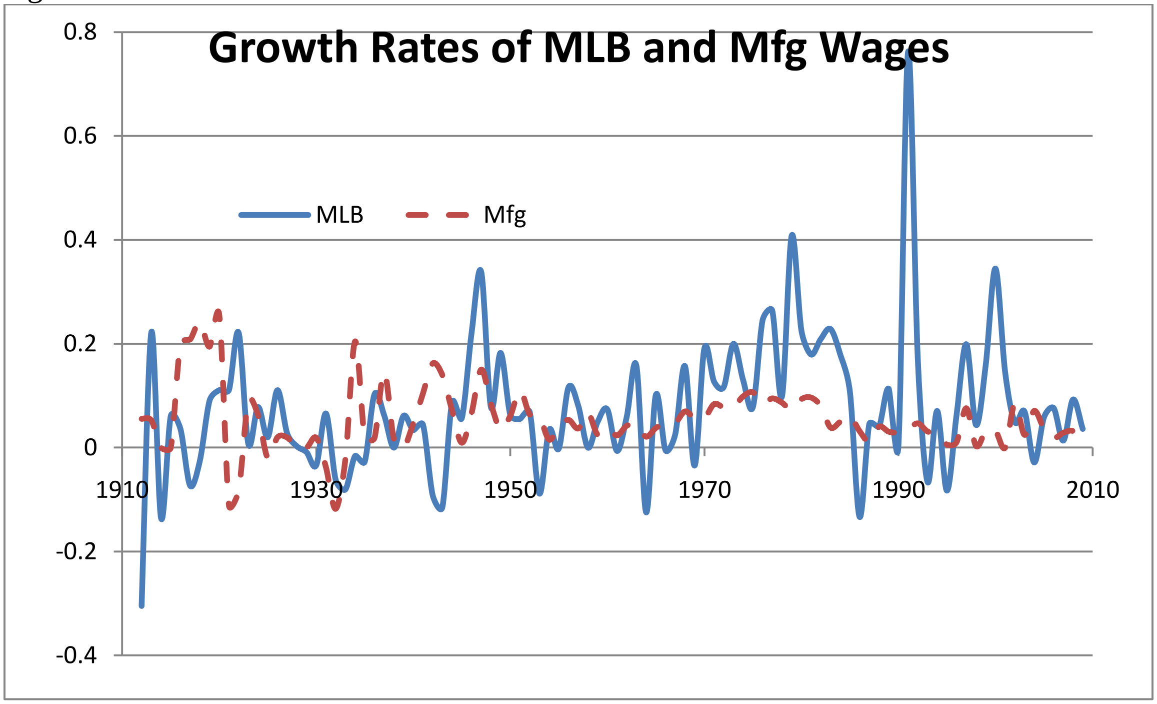


Figure 3


Growth Rates of MLB and Mfg Wages
MLB
Mfg
0.8
0.6
0.4
0.2
0
-0.2
-0.4
1910
1930
1950
1970
1990
2010

References

Baseball-Reference.com website, accessed January 2010, http://www.baseball-reference.com/

Burger, John D. and Stephen J.K. Walters, "Market Size, Pay, and Performance: A General Model and Application to Major League Baseball," *Journal of Sports Economics* 4(2), pp 108-125 (2003).

Cramer, Richard Ben, *Joe DiMaggio: The Hero's Life,* New York: Simon and Schuster, 2000.

Fort, Rodney, "Pay and Performance: Is The Field of Dreams Barren?" In *Diamonds Are Forever: The Business of Baseball*, ed. Paul M. Sommers, Washington, DC: Brookings (1992), 134-62.

Frank, R. H., "Are workers paid their marginal products?" *American Economic Review* 74, no. 4 (Sep 1984), 549-71.

Hamilton, James, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357-384, 1989.

Hamilton, James, *Time Series Analysis*, Princeton University Press, 1994.

Haupert, Michael J., "Player Pay and Productivity in the Reserve Clause and Collusion Eras," *Nine: A Journal of Baseball History and Social Policy Perspectives* (Fall 2009), pp 63-85.

Hoaglin, David C. and Paul F. Velleman, "A Critical Look at Some Analyses of Major League Baseball Salaries," *The American Statistician* 49:3 (August 1995), pp 277-85.

Kahn, Lawrence, 1993, "Free Agency, Long-term Contracts and Compensation in Major League Baseball: Estimates from Panel Data," *The Review of Economics and Statistics* 75:1, pp 157-64.

Kahn, Lawrence M., "The Sports Business as a Labor Market Laboratory," *Journal of Economic Perspectives* 14 (3), summer 2000, pp 75-94.

Kim, Chang-Jin and Charles R. Nelson, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press, 1999.

Krautmann, Anthony C., "What's wrong with Scully-estimates of a player's marginal revenue product?," *Economic Inquiry* 37 (2), April 1999, pp 369-81.

Krautmann, Anthony, Elizabeth Gustafson, and Lawrence Hadley, "A Note on the Structural Stability of Salary Equations: Major League Baseball Pitchers," *Journal of Sports Economics*, 4 no. 1, February 2003, pp 56-63.

Krautmann, Anthony C. and Margaret Oppenheimer, "Contract Length and the Return to Performance in Major League Baseball," *Journal of Sports Economics*, (February 2002) 3:1, pp 6-17.

Lowenfish, Lee, *The Imperfect Diamond: Baseball's Labor Wars*, New York: Da Capo Press, 1980.

MacDonald, Don N. and Morgan O. Reynolds, "Are Baseball Players Paid Their Marginal Products?" *Managerial and Decision Economics* 15, no. 5 (Sep-Oct 1994), 443-457.

Medoff, M. H., "On monopsonistic exploitation in professional baseball," *Quarterly Review of Economics and Business* 16 (1976), 113-121.

National Baseball Hall of Fame, Transaction Card database.

Schechter, Gabe, "8th-Inning Relievers: What Are They Good For?" presentation at SABR National Conference, Seattle, WA, 2006

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

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

Seymour, Harold, *Baseball: The Early Years,* 1960, New York and Oxford: Oxford University Press, 1989.

Zimbalist, Andrew, "Salaries and performance: beyond the Scully model. In *Diamonds Are Forever: The Business of Baseball*, ed. Paul M. Sommers, Washington, DC: Brookings (1992a), 109-33.

Zimbalist, Andrew, *Baseball and Billions*, New York: Basic Books, 1992b.