5-4-2010

Are Major League Baseball Starting Pitchers Compensated for Stadium Risk?

Joseph Glatman Zaretsky
Macalester College, jglatmanzaretsky@macalester.edu

Are Major League Baseball Starting Pitchers Compensated for Stadium Risk?

Joseph Glatman Zaretsky
Economic Senior Honors Thesis
May 3, 2010

Adviser: Professor Karine Moe
Readers: Professor Pete Ferderer and Professor Daniel Gilbert

Abstract: This paper examines the significance of stadium effects on the determination of starting pitcher salaries. It models stadium effect first under the assumption of perfect certainty, and then includes risk through uncertainty. Using starting pitchers' statistics between 1990 and 2008, this paper determines that the stadium effect is not significant in the model with perfect certainty, but becomes significant when uncertainty (risk) is introduced. An unexpected result of the test shows, however, that there is a fundamental difference between the American League, where the stadium effect is significant, and the National League, where it is insignificant in both models.

Acknowledgements:

I would like to extend my gratitude to the numerous individuals who have provided extensive guidance and assistance throughout the course of this thesis. Thank you to my adviser, Professor Karine Moe, for agreeing to assist with the project. I sincerely appreciate the countless times you met with me to read through all of the revisions, and for helping me find solutions to every problem that arose. I would also like to thank my readers, Professor Pete Ferderer and Professor Daniel Gilbert, both for the various comments I received on the project, and also for agreeing to be on my committee. Additionally, I would like to thank Professor Raymond Robertson. Not only did he teach me everything I needed to know for the econometric analysis, but also, without his confidence in me to complete this project, I likely would not have made it this far. I would also like to extend thanks to Trevor McCalmont, who worked with me through some of the complicated steps in the derivation of the theory section. I would like to also recognize my fellow honors seminar students for their daily humor and support. Lastly, I would like to extend a special thanks to my family and friends for supporting and encouraging me throughout the entire process.

Introduction

The Major League Baseball labor market provides a unique opportunity to test labor economic theories on the determinants of compensation, and the findings of these tests are easily transferable to other labor markets. Player performance statistics are publicly available, as is player compensation. This public access to productivity and pay data creates a distinct advantage over other labor markets, where productivity is usually difficult to measure and compensation is often private.

Typical econometric papers that study Major League Baseball have focused on the determination of player salaries through understanding each player's respective marginal revenue product (MRP). The MRPs for Major League Baseball players are calculated through the various performance statistics that the players accumulate over their careers. This paper broadens the analysis of salary determination by estimating the effect of the differences in stadiums on player salaries. With new stadiums opening around the country, there is much discussion about the effect certain types of ballparks have on the statistics of players and teams.[1] In particular, many of the new ballparks are said to be "hitter's parks", often because of short home run fences or high elevation, both of which give a distinct advantage to hitters over pitchers.[2] Pitcher performance in these stadiums would likely suffer and may reduce the pitcher's marketability to move to other

[1] There have not been any econometric studies. The discussion lies between journalists (e.g. ESPN) and statisticians within Major League Baseball.

[2] For example, stadiums like Coors Field are built high above sea level, giving the ball added lift when traveling in the air, usually leading to more home runs.

teams in the future. With this knowledge, do teams in a hitter's park have to compensate a pitcher for the expected reduction in his performance?[3]

The current literature misses the opportunity to test whether players are compensated for the inherent risks they undertake when they choose a particular home stadium. This paper includes both the known and the expected differences between stadium effects and argues that the current models to determine player salaries cannot fully explain the compensation of pitchers. This paper expands the literature in two ways; it improves the explanatory power of the traditional MRP estimations, while it simultaneously tests whether pitchers receive compensation for the risks inherent in playing in a hitter's park.

The remainder of this paper consists of five sections. Section 1 reviews the previous literature that influences this paper. Section 2 explains the econometric theory behind this question. Section 3 then summarizes the data that are used in the empirical work. Section 4 presents the analysis of the data, including tests to ensure there are no estimation issues, and regression results for both the perfect certainty model and the uncertainty model. Section 5 concludes this study and provides ideas for future research.

1) Literature Review

Most research on player salaries in Major League Baseball tests the theoretical argument that players, like other workers, are paid according to their MRP. Rottenberg (1956), who was one of the first to conduct a study regarding player MRP, concluded that many players were paid below their MRP. He argued that teams exploited the players during

[3] The stadium effect comes from a team's home ballpark only, as teams are unable to control the stadiums they play in outside of the home park.

the bargaining process, in the era before free agency. During this period, players could only be free agents if their teams released them. Once a player was under contract however, a constant renewal clause allowed teams to use the player at the same salary at the end of each contract period.[4]

Scully (1974) econometrically measured baseball players' values to determine how overpaid or underpaid players are. He estimated player's MRPs, and from that produced a model to determine projected salary of the players. The model incorporated equations for three variables: team revenue, team winning percentage, and team performance. Scully estimated that teams paid players only 10 to 20 percent of their MRP, likely caused by the player's lack of bargaining power.[5] Scully's study, however, was conducted before free agency and thus does not take all factors that we see today into account.

Major League Baseball established free agency in 1975, and since then, the MRP continues to be the chosen method to determine the value of players in the academic literature. The recent literature, however, focus on position players because their performance is fairly easy to compute.[6] Fewer studies have focused on pitchers. A pitcher's individual performance is not as easily determined because many times it is a reflection of the skill of the fielders behind the pitcher. One specific question that has not been examined is whether stadiums affect pitcher salaries. Variation in a pitcher's home

---

[4] Free agency allows players to negotiate with teams for new contracts once their rookie contract expires. Originally, there was a renewal clause in almost every contract, which allowed teams to renew a players' contract, severely limiting the bargaining power of the player. After a 1975 lawsuit, this rule was changed.

[5] The model also includes standard performance variables such as ERA, Innings Pitched, and Strikeouts.

[6] The previous literature has also concluded that statistics are in fact the driving force for salary offers.

stadium might affect his statistics and, to the extent that pay is based on statistics, his salary.[7]

As free agency established itself in the baseball industry, newer studies began to find that players were paid closer to their MRP. This is understandable, as once a player can negotiate with any team and sell his services, he should sign with the team that pays him closest to his MRP. Kahn (1993) updated Scully's (1974) original research and used career statistics, as opposed to season-by-season statistics, to determine the performance factor for individual players. He argued that career statistics are better predictors of future production and thus better represents what a team examines when it determines player values.[8] Kahn (1993) found that players were now paid right near their MRP.

Krautmann et al. (2003) separated pitchers from other players, and additionally divided pitchers into starters, long-relievers (those who enter the game after the starter, and before the stoppers), and stoppers (those who finish the game, typically the final inning). The study determined that performance variables for the starters vary more than for relievers and stoppers, and therefore may have a more significant effect on the determination of a starter's salary than for a reliever's performance.[9] This result suggests that starters may be likely to feel the stadium effects more than relievers and stoppers. If relievers and stoppers are typically strikeout pitchers, the difference between stadiums

---

[7] A pitcher plays half of his games in his team's home stadium each season.

[8] Kahn (1993) did not, however, use counting statistics such as strikeouts and innings pitched, and instead only used career ratios, such as ERA and winning percentage

[9] Performance variables include strikeouts, walks, ERA, innings pitched, and type of pitcher. The variance of pitcher-type is greater in starting pitchers than others. Starters can be strikeout pitchers, pitchers who pitch to contact, or a combination of the two depending on the situation. Relievers and stoppers, however, tend to fall mainly in the strikeout category.

should not affect their statistics. Starters, however, vary in type. This would therefore cause a starter's statistics to reflect the stadium effect more than relievers and stoppers. Starting pitchers allow the ball to be put in play more often than relievers, providing more chances for the stadium to have an effect on the pitcher.

The dearth of studies in stadium effects leaves a hole in the literature. Goodman and McAndrew (1993) offered the only study of stadium effects on player performance. They tested how the performance of hitters varied when playing on either grass or Astroturf, a surface typically found in domed stadiums. Using data from the late 1980s, they concluded that dome stadiums with Astroturf were more conducive to offensive players.[10] This result is now outdated however, because most of the domed stadiums studied are no longer in use and Astroturf is no longer the surface used in the new domes.

Few papers have addressed the role risk plays in salary determination in Major League Baseball. Krautmann and Novak (2004) examine whether injury risks to MLB catchers and second basemen had an effect on their salaries. They believed that compensation might also have an effect on which team they choose to play for. The results, however, found that there is no compensation for risk. They also tested compensating wage differentials by examining a player's desire to take a lower salary in exchange for more desirable job conditions, specifically, living in cities that are located in better climate zones. The authors concluded that players did indeed take a lesser salary to accommodate their desire for a better climate, therefore allowing certain small-market teams in the regions with the best climate to sign players otherwise out of reach. The

[10] Statistics such as batting average and homeruns hit were higher in the fields with Astroturf.

paper only analyzed hitters, however, once again citing the difficulty in evaluating pitchers as the reason.

Substantial literature on risk and wage theory also exists outside of Major League Baseball. One specific type of risk that is relevant to this thesis is earnings risk in a labor market. The initial studies of earnings risk focus on employment instability. Feinberg (1981) used a six-year survey of 1,419 families and determined that the variation found in income over time is in fact correlated with salary determination.[11] Workers of equal characteristics will receive different salaries if one has a higher degree of employment risk. That worker will receive compensation for earnings risk in the form of a higher total salary. Should a pitcher believe his statistics will get worse in a hitter's stadium, he may deduce that there is instability in his expected future income and therefore require the risk compensation.

Rosen (1986) surveys the literature on worker compensation models. The paper is not an empirical paper. He instead defended his theories by referencing past papers that studied specific labor markets. Citing various studies, he concludes that there is evidence that firms typically compensate workers for earnings risk. For example, he discusses one model that explains that firms will compensate workers at a risk premium if the specific occupation does not allow laborers to work the desired number of hours at the given hourly wage. This risk is similar to the stadium risk, as teams that play in a hitter's stadium understand that its environment is not ideal for pitchers, and must therefore compensate them for the less-desirable situation.

[11] The study uses mean wage as the dependent variable, while its control variables include: age, race, sex, education, occupation, and union. Instability is measured by variation in annual income over six years.

Another form of uncertainty with earnings risk is the question about future wages; whether workers receive stable wage increases over time, or if future wages are unknown. Mcgoldrick and Robst (1996) found that industries in which future earnings are unstable tend to compensate workers for that uncertainty. Current wage offers in those industries are therefore greater than in occupations in which future earnings are more certain.[12]

The Major League Baseball labor market is important to study, in part, because it is a well-publicized, unionized market. Many of the risk compensation studies do not discuss the inherent differences between union and non-union industries. They include unions as control variables, but do not examine why union workers are treated differently than non-union workers. Moore (1995) compared risk compensation between union and non-union markets. He examined industries where workers risk losing hours because the demand for output is unpredictable. Moore used a risk-union interaction term, along with the typical variables included to establish workers' salary. He determined that union members are paid a higher percentage increase in their salary than non-union workers in similar job markets, with every added unit of risk. This result explains that earnings risk is most recognized in unionized labor markets. Therefore the examination of the Major League Baseball labor market may lend support to the study of other unionized industries. This is an important distinction because the quality of salary bargaining in Major League Baseball is much more comparable to other unionized markets instead of non-union markets.

---

[12] The study uses mean industry wage as the dependent variable, while its control variables include age, race, sex, education, occupation, and union. "Income risk is defined as the standard deviation of residual earnings from individual earnings functions estimated over time and time squared." Mcgoldrick and Robst (1996)

It is unclear whether these compensation theories hold for pitchers in Major League Baseball. This paper tests these theories for individual pitchers by estimating stadium effects on pitcher salary determination. It first models stadium effects under perfect certainty. In this case, both the pitcher and team know what the effect is with no uncertainty. The second model allows for the uncertainty of risk, which occurs when the exact effect of the stadium is unknown. The examination of the uncertainty model provides a conclusive result that answers the question: if a pitcher's statistics are at risk of getting worse because of where he plays, will that affect the determination of his salary and lead to compensation?

2) Theory

The goal of a professional pitcher, as is the case with other labor service suppliers, is to maximize utility. The main component of the utility function is consumption, which in this study is assumed to be a function of income (I), so that: $\partial U/ \partial I > 0$.

I derive the formal model in two situations. The first is the case of perfect certainty about the effect of a hitter's stadium on each pitcher's statistics, and the second is to include uncertainty. When there is perfect certainty in the model, the measure of change in a pitcher's statistics is not a risk variable because the effect that each stadium has on the pitcher's statistics is known. The model with uncertainty involves risk because the exact effect of each stadium is unknown.

Model with Perfect Certainty

The model assumes that a pitcher's performance and statistics determine his salary. It is a two period model, where the pitcher maximizes expected lifetime income by choosing between two offers, one from a team with a hitter's park home stadium, and the other with a neutral stadium. The options are summarized as follows:

$$V_1 = U(Y_{11}) + \beta U(Y_{12}(x(g))) \quad (1)$$
$$V_2 = U(Y_{21}) + \beta U(Y_{22}(x(g))) \quad (2)$$

where V_1 is the summation of current and discounted future income (utility) when playing in the hitter's stadium, and V_2 is the sum in a neutral stadium. Y_{11} is the current salary the hitter's stadium team offers to the pitcher, and is a function of player statistics to date, while Y_{12} is the future salary offer. Y_{12} is a function of the statistics of the pitcher from the first period (x), which is in turn a function of the stadium effect (g). The same is true in the neutral stadium.[13]

The decision rule for the pitcher will be to choose the stadium option that provides the highest salary, as more income leads to a higher utility. He will therefore only consider playing in the hitter's stadium if the sum of his salaries is greater than or equal to the sum from the neutral stadium ($V_1 \geq V_2$). The model assumes that a pitcher's salary depends on his performance. We therefore may claim that V_1 and V_2 are known in the current time period. This is true because in an ideal environment with perfect certainty, teams know exactly how the pitcher will perform in both stadiums. Therefore, his current and future salaries are already known. In the hitter's stadium, the current and future salaries are still known because the effects of the stadium are also known for that stadium.

With these assumptions, we may now solve for Y_{11} and derive the equation to understand the relationship between the stadium effect and salary in an ideal environment

[13] Y_{22} is a function of the statistics from the first period in the neutral stadium, which is in turn a function of the neutral stadium effect.

with perfect certainty. We begin by setting V_1 equal to V_2, as that is the lowest value of V_1 that the pitcher will accept to play in the hitter's ballpark, and solve for $U(Y_{11})$.

$$U(Y_{11}) = U(Y_{21}) + \beta[U(Y_{22}) - U(Y_{12})] \quad (4)$$

The goal in the following derivation is to find $\partial Y_{11}/ \partial g$.

All of the steps in the derivation can be found in the Appendix. The result of the derivation is as follows:

$$\partial Y_{11}/ \partial g = \beta[\partial Y_{22}/ \partial x * \partial x/ \partial g - \partial Y_{12}/ \partial x * \partial x/ \partial g] \quad (5)$$

If all teams know the exact effect of each stadium, then in the future, teams that offer a pitcher coming from a hitter's stadium a contract are able to separate the effect of the stadium from his statistics and offer him a contract equal to his worth prior to the stadium.[14] Therefore, in equation (6), it can be considered that g is equal to zero in both the neutral stadium and the hitter's stadium sections. If $g = 0$, then the equation overall will also equal zero, so that, $\partial Y_{11}/ \partial g = 0$.

The hypothesis for the model with perfect certainty therefore, is that a pitcher will not in fact receive any compensation for stadium risk.

Model with Uncertainty

In the model with uncertainty, the stadium effects on statistics are not perfectly known. Along with the assumptions from the model with perfect certainty, this model also assumes that players are risk neutral. There may be a great effect, making one pitcher's statistics significantly worse, or a pitcher can perform better than the expected outcome. We begin again with the decision between V_1 and V_2.

$$V_1 = U(Y_{11}) + \beta EU(Y_{12}) \quad (6)$$

[14] Assuming the rest of his skills did not diminish.

$$V_2 = U(Y_{21}) + \beta EU(Y_{22}) \quad (7)$$
$$U(Y_{12}) = p(g)(Y_{12}^{Low}(x(g))) + (1\text{-}p(g))(Y_{12}^{High}(x(g))) \quad (8)$$

The uncertainty is found in Y_{12}.[15] There are now two outcomes that are possible when playing in the hitter's stadium. Either it will affect the pitcher's statistics by a large amount, leading to a low future salary offer, or there will not be a great effect on his statistics, and he will therefore receive a standard salary offer. These possibilities are labeled Y_{12}^{Low} and Y_{12}^{High}, where p(g) is the probability of a large effect, and (1- p(g)) is the probability of a small effect.

With this new assumption we use the same steps as the basic model to derive the final equation. The full mathematical derivation can be found in the Appendix. The result is as follows:

$$\partial Y_{11}/ \partial g = -[(p(g)(\partial Y_{12}^{L}/\partial x * \partial x/ \partial g * g) - (1 - p(g))(\partial Y_{12}^{H}/\partial x * \partial x/ \partial g * g))] \quad (9)$$

Once again, this result shows that the relationship between the type of stadium and Y_{11} is positive; therefore a pitcher should theoretically be compensated to play in a hitter's ballpark.

We therefore come to the final guiding equation:

$$Salary = \alpha_0 + \beta_1 Performance + \beta_2 Stadium\ effect + \beta_3 Misc + \varepsilon \quad (10)$$

where *Performance* contains the various performance variables that are taken into account to determine pitcher salary and *Misc* contains miscellaneous team variables that would affect the salary offered by the team's management.

---

[15] Both Y_{12}^{Low} and Y_{12}^{High} are functions of the statistics from the current period, which is again a function of the stadium effect.

3) Summary Statistics

In order to test both the perfect certainty and uncertainty models, two different measurements must be used. To test the perfect certainty model, this study uses the park factor variable as a measure for the stadium effect, where the park factor indexes the average number of runs scored in a stadium to the league average.[16] By measuring the specific effects of each stadium on the number of runs scored per game, the park factor proxies for the effect of the stadium on pitcher performance. The park factor variable works well as a proxy for the perfect certainty model because while the means of ERA vary across stadiums, there is no significant difference in the variance. Thus in an ideal environment, the stadium effect on a pitcher's ERA is known.

To test the uncertainty model, this study uses a dummy variable denoting whether or not a stadium is a domed stadium or an outdoor stadium. Using the test on the variance and mean of ERA in indoor and outdoor stadiums, this study determines that outdoor stadiums are riskier for pitchers than indoor stadiums overall. There is no difference in the means of ERA, but the variance is greater in outdoor stadiums. Therefore, one cannot determine the exact change in a pitcher's ERA between stadium types. These differing variances might be explained by factors such as the highly elevated

[16] ESPN calculates the Park Factor as: Park Factor = ((homeRS + homeRA)/(homeG)) / ((roadRS + roadRA)/(roadG)), where: homeRS=Runs scored at home, homeRA=Runs allowed at home, homeG=number of home games, roadRS=Runs scored on the road, roadRA=Runs allowed on the road, roadG=number of road games. The park factor is equal to the average number of runs scored in a stadium indexed to the league average. A neutral stadium has a value of 1, while a hitter's ballpark is any stadium with a value over 1. The variable measures the specific effects each individual stadium has on the amount of runs scored. In other words, if the pitcher's stadium has a park factor of 1.1, then presumably the pitcher will allow .1 more runs per game then in a typical stadium in the league (park factor is expressed in runs/game).

outdoor stadiums such as Coors Field, or the small dimensions found in many outdoor stadiums such as Fenway Park.[17]

1) Model with Perfect Certainty

The data collected for the model with perfect certainty (using the park factor variable) contain 3,719 observations that are season-by-season statistics of every pitcher that has started over 60% of the games they have appeared in from 1990-2008.[18] Major League Baseball recorded the statistics for each game and posted them to final box scores. At the end of each season, the statistics are added together to obtain the total season's statistics. These statistics were collected from both ESPN.com and Baseball-Reference.com. Revenue data are available through the financial reports of each team. Though the teams are privately owned, and standard regulations do not require the release of their financial data, each team reports its revenue, according to Major League Baseball's own policies. These reported revenues were available on Forbes.com, and the revenues were manually entered into the data set. Teams typically make salary data publicly available and were collected by looking at each starting pitcher's player pages on ESPN.com, and Baseball-Reference.com. The park factor variable was also found on ESPN.

Table I summarizes the final data and figure 1 presents histograms for the respective variables. The yearly values of salary range from $100,000 to $22,000,000. As expected, the histogram indicates a significant right skew seen in the salary data.

---

[17] Coors field is located in Colorado, while Fenway is located in Boston.

[18] This distinction was made because some starting pitchers will come in as relievers if necessary. If a pitcher is a reliever in more than 40% of their appearances, however, it can be inferred that teams are likely not paying them as a starting pitcher, but rather as a reliever or spot-starter.

Taking the natural log of salary corrects for the skew.[19] If not recorded, the salary for rookies and players recalled from the minor leagues is assumed to be the minimum salary for that year.[20]

The range for team revenue is also quite large, with a maximum of 327 million, and a minimum of 24.9 million. The mean is approximately 106 million. The variable shows a slight skew to the right, necessitating the natural log. The variable is divided by 1,000,000 for simplicity. Salary offers also typically change depending on the age of the player. The mean age is 27.8 years, with the youngest pitcher being 19 and the oldest 46. The team performance variables, fielding percentage and winning percentage, are both normally distributed as seen in their histograms. The fielding percentage variable does not vary much as the minimum is 97.2% and the maximum is 98.9%, with a mean of 98.2%.

The individual performance variables are all relatively normally distributed. The average number of games started per year is approximately 22, with a maximum of 37. Hits allowed during the season range from two to 284, with a mean of 138.4 and a standard deviation of 70.6, as some pitchers only started one game in a certain season. The means for walks and strikeouts are 46.7 and 93 respectively, with large ranges once again because of the varying number of games started. Strikeouts again show the different types of pitchers with a standard deviation of 59.71.

---

[19] This results in an almost normally distributed curve, however there is still a large extension on the lower end of the salaries because there are so many minimum salary contracts.

[20] Minimum salary ranges from $100,000 in 1990 to $390,000 in 2008.

Instead of including both wins and losses, I simply use the win-loss percentage, which has a mean of .479, showing that the average pitcher actually has a losing record.[21] The earned run average (ERA) has a range of zero to 43.2.[22] The average ERA is 4.9.

To control for free agency, a dummy variable is also included to indicate when a pitcher signed a new contract.[23] Unfortunately, news of every new contract signed is not readily available. This is evident by its small mean of 16.6%. The results of the study therefore may not fully capture every time a new salary is offered to a pitcher.

The final variable is the park factor. The variable is normally distributed and, as expected, has a mean of 1. The standard deviation is .114. If all of the stadiums were the same, then the same number of runs would presumably be scored in each stadium, leading the park factor to equal one for each stadium. Since all stadiums are not the same, it is simply a measure of how many more runs, on average, are scored at one park compared to the other parks. Therefore, the mean of one was expected, and the standard deviation of .114 is beneficial for this study, as it is fairly large for a variable whose minimum and maximum is .606 and 1.412, respectively.[24]

2) Uncertainty Model with Risk

The data used to examine the uncertainty model contain 24,268 observations that are the game's logs for every game pitched by every starting pitcher between 2004 and 2008,

---

[21] Starting pitchers do not record wins and losses for every game in which they participate. Sometimes relievers get the decision depending on the situation.

[22] One pitcher had an ERA of 99.99, however he was removed because he was an outlier that had not recorded any innings pitched.

[23] An observation with a value of 1 indicates a year the player signed a new contract.

[24] The stadiums with the minimum and maximum park factors are Petco Park and Coors Field respectively.

plus his yearly salary, and the yearly team performance variables. The data for this model are different from the perfect certainty model so that we can investigate the specific differences that arise when pitching in a domed stadium compared to an outdoor stadium on a game-by-game basis.

The editors from Baseball-Reference.com entered these statistics from the box scores into a comprehensive dataset, and allowed its use for this study. I calculated some statistics, such as game ERA, and innings pitched per game, using the given data, and manually entered them into the data set. The remaining data, such as team revenue and salaries, are collected from the same sources as the data from the perfect certainty model.

One problem with the data arose in the unavailability of information regarding retractable roofs. There is no available record that shows which specific games were played with the stadium roof open, and which were played with it closed. There are also games during the season in which the roof is closed or opened mid-game. These games are only found in a small percentage of the data set. The stadiums that have retractable roofs are: Arizona, Houston, Milwaukee, Seattle, and Toronto. We considered these games to be indoor games, as any time there is a weather factor that may affect a pitcher's statistics, the roof is typically closed.

Table II summarizes the data used in the uncertainty estimation and figure 2 presents the histograms for the respective variables. Before changing the retractable roof stadiums to indoor, approximately 90% of all games were played outdoors. Once changing the variable to include retractable roof stadiums, the difference between the two possibilities becomes more useful. The mean now shows that 76.2% of the games were played outdoors.

The salary variable is the dependent variable. The yearly values range from $300,000 to $22,500,000. There is again a significant right skew in the data when viewing the histogram, necessitating the natural log.[25] If unreported, the salary for rookies and players recalled from the minor leagues is assumed to be the minimum salary for that year.

The variable for team revenue has a maximum of $327 million, while the minimum is $80 million, and the mean is $156 million. The variable is also normally distributed. There is a clear distinction between the small-market teams that may be less willing to compensate their pitchers, and the large-market teams that typically have more money to spend. Small-market and large-market teams typically do not change from year-to-year, so it is usually known which teams have the most money to spend. The revenue variable is divided by 1,000,000 for simplicity.

The team performance variables, fielding percentage and winning percentage, are both normally distributed. This was expected because there are typically only a few elite teams, and a few very bad teams, with all of the others in between. Even with the difference between elite and poor teams, the fielding percentage does not vary much as the minimum is 97.7% and the maximum is 98.9%, with a mean of 98.3%. Team winning percentage ranges significantly with the minimum of .315 and the maximum of .648, with the mean predictably falling at .499.

The five individual player performance variables are all relatively normally distributed. Innings pitched per game ranges from one third of an inning to 10 innings,

[25] This results in an almost normally distributed curve, however there is still a large extension on the lower end of the salaries because of so many minimum salary contracts, as can be seen in the first histogram.

with a mean of 5.859 and a standard deviation of 1.52. Hits and walks allowed both have minimums of zero, and maximums of 15 and 10 respectively. Strikeouts also vary, as the range is from zero to 18, with a mean of approximately four. The fairly large standard deviation of 2.39 shows how some pitchers are strikeout-pitchers, while others are not. The most interesting variable is ERA, in which the range is from zero to 270. The range is large because some pitchers allow many runs in just one third of an inning, before being replaced. With the maximum value so high, and with so many games in which zero earned runs were allowed, the histogram returns a slight skew to the right, but is still relatively normally distributed. This range also explains the rather large standard deviation of 8.38.

4) Analysis

Perfect Certainty Model Results

Table IV reports the final regression results.[26] Regression (i) shows the main regression using the full data set. The park factor variable returned a positive coefficient, indicating that if the park factor increases by one run-per-game, the salary of that pitcher should increase by 17.3%. The coefficient, however, is statistically insignificant, as can be seen by the absolute value of the t-statistic equal to 1.42. Therefore, this result follows the original hypothesis that players are not compensated in the perfect certainty model. Due to the fact that park factor is an imperfect proxy for the perfect certainty model, the coefficient returned a positive value instead of simply zero.

The team revenue variable returned a positive and significant coefficient with a value of 0.754, and a t-statistic of 23.98. This result is not surprising, as team revenues

[26] See Appendix B for estimation issues.

range widely and teams with higher revenues have much larger payrolls. This is evident in the yearly revenue reports, as the large-market teams such as the New York Yankees and Mets annually have the highest payrolls. Small-market teams such as the Oakland Athletics and Pittsburgh Pirates regularly reside near the bottom of the list. The other team variable, fielding percentage, returned a surprising coefficient, as it was much larger than any other variable and significant at the 1% level. The coefficient returned a value of 14.131. This may be a result of the more elite and richer teams possessing fielders who are superior to the less wealthy teams that offer pitcher smaller contracts.

The coefficients on the age and age^2 variables are positive and negative respectively, and both are significant. This result is consistent with expectations, as teams should offer a pitcher at a young age a higher salary as he improves. Once he reaches a certain point however, his potential improvement diminishes, and his salary offers do not increase nearly as much.

The signed contract variable, however, does not return the expected result. Theory hypothesizes that, each time a player signs a new contract, his salary offer typically increases. The significant results, however, show that the average time a player signs a new contract, his salary decreases by 30.2%. This is highly unexpected, and the reason for this is unknown. One explanation however, could be that the limited information regarding when players signed a new contract left the variable incomplete. Should all of the new contracts and contract extensions be made readily available, the result of the regression may return a coefficient closer to expectations. Another possibility is that the good players sign long-term contracts early in their careers. Therefore, the players signing new contracts most often are those who are not as skilled.

Therefore, as they continue to sign more contracts, their offers lessen because they are not the good players in the league. The other coefficients do not change much if the signed contract variable is removed.

Some of the results for the performance variables were unexpected. The variables that return expected results are strikeouts per game (SO/GM), games started (GS), innings pitched per game (IP/GM), walks per game (BB/GM), win-loss percentage (W-L%), and opponent on base percentage plus slugging percentage (OPS). Walks per game, win-loss percentage, and on base percentage plus slugging percentage, all return insignificant coefficients. The coefficient on strikeouts is a positive value of 0.179, and is significant at a 1% level with a t-statistic of 10.12. The value of the coefficient is consistent with expectations, as it usually garners a lot of attention when evaluating the skills of a pitcher in the previous literature.

The performance variables that do not follow the initial hypothesis are hits per game (H/GM) and ERA. Both hits and ERA returned positive coefficients significant at a 5% and 1% level respectively. This result implies that with each added hit and each added run (per nine innings) allowed by the pitcher, his salary should increase by 7.6% and 5% respectively. This raises concern, as the more hits and runs a pitcher allows should not increase the amount that he is paid. This result suggests that the model may be wrongly specified, though it is consistent with the theories used in the previous literature. One explanation for this inconsistency may be that the pitchers with the highest salaries typically pitch more innings than other pitchers. Therefore there is a greater opportunity for those high paid pitchers to allow more total hits than the other starters. The ERA variable still leaves questions, as it is hypothesized that the higher paid pitchers have lower ERAs.

Perfect Certainty Model Robustness

The surprising results must be verified with robustness tests. In order to ensure that the park factor variable is not the cause for the strange results, regression (ii) shows the results with the variable omitted. No coefficients changed by any significant degree.

The next test to better understand the results is to split the data between the National and American Leagues, the two leagues in Major League Baseball. Regressions (iii) and (iv) report the results, with the third regression using the American League data, and the fourth regression using the National League data. The results between the two regressions show a clear distinction between the leagues. The results of the American League regression return the same signs on each coefficient when compared to the original regression, though some variables lost their significance.[27] The one main difference, however, is that the park factor variable is positive and significant at the 5% level. The results indicate that a pitcher's salary increases by 48.5% with an increase of a full run-per-game in park factor. This coefficient is much larger than that from the original regression, while the values of the other variables do not change very much, other than team fielding, which both lost its significance and saw its coefficient change to 0.689. As we can see in the summary statistics, park factor never actually increases by a full run; therefore the likely increase in pitcher's salary will be much less than 48.5%. Instead, if the park factor increases by one standard deviation, the estimation implies that

[27] Innings pitched, hits, and team fielding.

a pitcher's salary would increase by 6.3%. Therefore, the results from the American League does not follow the original hypothesis.

The National League regression also changes very little in the performance and team variables,[28] but the park factor variable becomes very small, negative, and insignificant. This result shows that there is a clear difference in the way the two leagues value their pitchers and determine their salaries.

The biggest reason for the significance seen in the park factor variable in the American League and not the National League, is likely attributed to the designated hitter (DH) rule. The American League uses designated hitters that bat for the pitchers, while the National League must have the pitchers hit, therefore making offense in the American League much stronger. To attempt to better understand the statistical difference between the two leagues, tables V and VI show the summary statistics of the leagues. The ERA variable is much larger in the American League, and therefore may be the cause for the significant result of the park factor variable. The mean park factor variable in the American League is 1.011, while the mean in the National League is .992. This difference may account for some of the results as well, though it certainly would not account for all of it.

The final robustness test was to use only the pitchers where the signed contract variable is available. This was done to see if the ERA and hits variables might change when the data set is minimized to players where every variable is accounted for. The variables did in fact change, as ERA remained positive but became insignificant, and hits

[28] Only hits and team fielding variables become insignificant, and OPS become significant.

became negative and insignificant. This shows that the signed contract variable may be quite important for this study, as it began to correct the unexpected results.[29]

These robustness checks confirm the results of the performance variables, but introduce new questions regarding the difference between the American and National Leagues. It is clear that the results from the National League regarding the park factor variable overtake those from the American League in the initial regression.

Uncertainty Model Results

Table VIII reports the final regression results.[30] Regression (i) shows the main regression using the full data set. As hypothesized, the stadium type variable returned a positive coefficient with significance at the 1% level. The coefficient shows that by playing in an outdoor stadium, a pitcher's salary is likely to increase by approximately 6.8%.

The team revenue variable returned a positive and significant coefficient at the 1% level. The value of the coefficient explains that an increase in $1,000,000 of team revenue leads to a .42% increase in salary. The other team variables, winning percentage and fielding percentage, both returned insignificant results. The coefficient for winning percentage was negative, but insignificant.

The strikeouts variable returned positive and significant results at the 5% level, as was expected by the theory. The number of walks allowed by a pitcher, as well as innings pitched also followed the hypothesis, as the coefficients were negative and positive respectively and significant at the 1% level. Innings pitched has the largest

---

[29] The same test was also run with all of the statistics lagged to a year prior to the signed contract, but there was no significant difference between the results before and after lagging.

[30] See Appendix C for estimation issues.

effect on salary for any of the performance variables, as each added inning pitched *per game* is expected to increase salary by approximately 17.7%. This shows that the durability of a pitcher is relevant to teams as they determine his salary. Some of the results for the performance variables were unexpected, as in the perfect certainty model. The three variables were ERA, total hits allowed per game, and team winning percentage. Similar to the perfect certainty model, hits and ERA show positive coefficients, and were significant at a 1% level. This raises concern, as the more hits and runs a pitcher allows should not increase the amount that he is paid. The winning percentage was negative but insignificant. It was expected that if the team's winning percentage increased, the players would be paid more because they helped the team improve, but the results show the opposite. These results suggest that there may be a variable omitted that would help control for these unexpected performance factors.

Uncertainty Model Robustness

To test for robustness, I split the data into two sets: one for the American League, and the second for the National League.[31] The regression results from both leagues are reported in table VIII. Regression (ii) shows the results from the National League. The results for many of the variables remain unchanged from the main regression. The walks variable loses its significance, but retains it negative sign. The winning percentage variable becomes significant, but remained negative, which is a concerning result. The variable in question, stadium type also changed, becoming both negative and insignificant. Though this result is opposite from the initial regression, it may be

[31] Both data sets were once again tested for all of the potential estimation issues, and the same issues were discovered and corrected for in each set.

explained by the lack of offensive production in the National League, therefore, taking away the significant variance that was initially found.

The regression results using American League data support this hypothesis. Regression (iii) reports the results. As expected, the American League returned a positive and significant result on the stadium type variable. We see that in the American League, players are paid approximately 14.46% more if they typically play in an outdoor stadium as opposed to an indoor stadium. All other variables have the same signs and significance except for the winning percentage, which became significant. The coefficients themselves had only minor changes.

Surprisingly, there are more outdoor stadiums in the National League, as can be seen in tables IX and X respectively. The difference in compensation then may be a reflection of the better offense in the American League, or the fact that the American League has more elite teams.[32]

With these checks, it can be inferred that the original results are indeed acceptable, even though the National League actually diminishes the significance of that result.

5) Conclusion

This study examined whether or not Major League Baseball teams take the risk of pitching in a hitter's ballpark into account when determining the salary offers for starting pitchers. The results of this study showed that the risk of playing in an outdoor stadium as opposed to an indoor stadium does indeed have a significant effect on a pitcher's salary when examining the entire league. An unexpected result of the study, however, was the fundamental difference between the American League, where the risk of playing

[32] Team winning percentage is higher in the American League when compared to the National League.

in an outdoor stadium seems to be significant, and the National League, where the coefficient is not statistically different from zero.

The other unexpected results from the test, such as the opposite signs on various performance coefficients leads me to believe that there is either a variable missing, or specified incorrectly. There may be other variables considered by teams that are not as clear or known as these performance variables. The adjusted R^2 of the uncertainty model is approximately 0.15. Therefore other variables may be missing from this study and are unknown to this study or the previous literature. Certain variables that have recently been introduced to the baseball industry may provide a better method to evaluate the skills of pitchers. One specific variable is the Fielder Independent Pitching (FIP) statistic.[33] This variable calculates a pitcher's ERA independent from any situations where team fielding is involved, and only focuses on those statistics for which a pitcher is solely responsible. This statistic is a recent discovery and has not been made readily available for collection for many of the years studied in this paper.

The examination of the model with perfect certainty also returns both exciting and unexpected results. The park factor variable is not significant in the general regression, as was expected in the theory. Once again, however, the American League robustness check returns a positive and significant coefficient, while the National League was negative and insignificant. The R^2 value of each regression remains close to .599 however, which is a fairly strong value. Therefore, the likelihood of an omitted variable bias diminishes in this regression.

[33] FIP = (HR*13 + (BB + HBP – IBB)*3 – K*2)/IP

Further analysis into this subject could help teams determine whether or not certain salary offers are fair for both sides. If the data regarding signed contracts were fully collected, the regressions may be able to return more complete results. This seems clear, as the robustness test showed the potential effect that a fully collected signed contract variable may have. This study may be improved further by indicating when players were eligible for salary arbitration, and when they were signed away from their rookie contracts. This would help differentiate the rookie contracts and determine which were paid as rookies, and which were paid as top draft picks and given higher starting salaries.[34]

This study can also be implemented in other industries other than athletics, where there are similar environmental constraints on the performance of the worker.

[34] Not all rookies are treated the same in salary negotiations.

Table I. Perfect Certainty Model Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Salary	2,202,672	3,106,058	100,000	22,000,000
Tteam Revenue	106,052,500	54,297,250	24,900,000	327,000,000
Games Started	21.856	10.697	1.000	37.000
Innings Pitched	135.426	72.111	0.000	271.200
Hits	138.403	70.654	2.000	284.000
Earned Run Average	4.895	2.856	0.000	99.990
Walks	46.778	25.778	0.000	152.000
Strikeouts	93.005	59.724	0.000	372.000
Win-Loss Percentage	0.479	0.195	0.000	1.000
Team Fielding Percentage	0.982	0.003	0.972	0.989
Age	27.787	4.531	19.000	46.000
On Base Plus Slugging	0.777	0.128	0.340	2.333
Signed Contract	0.166	0.372	0.000	1.000
Park Factor	1.005	0.114	0.606	1.412

Notes: 3,082 Observations

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Oct. 2009.

Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Oct. 2009.

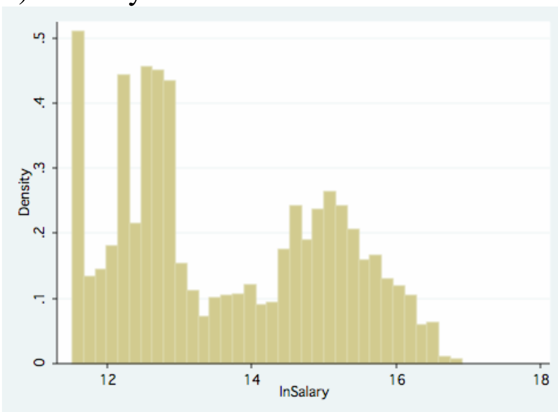
Player Index. The Baseball Cube. Major League Baseball. 27 Oct. 2009.

The Business of Baseball. Forbes. Major League Baseball. 27 Oct. 2009.

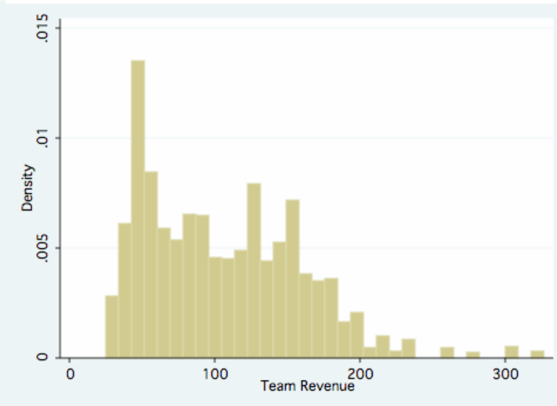
MLB Franchise Revenues and Expenses. Baseballchronology.com. http://www.baseballchronology.com/Baseball/Years/2005/Income_Expenses.asp. 27 Oct. 2009.

Figure 1. Histograms

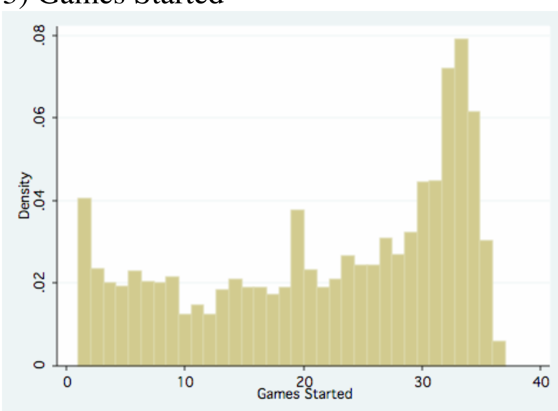
1) lnSalary



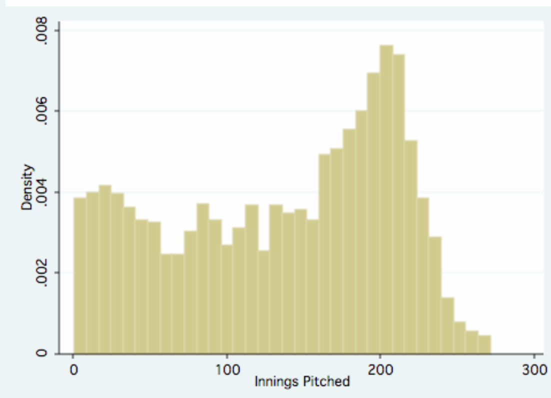
2) Team Revenue



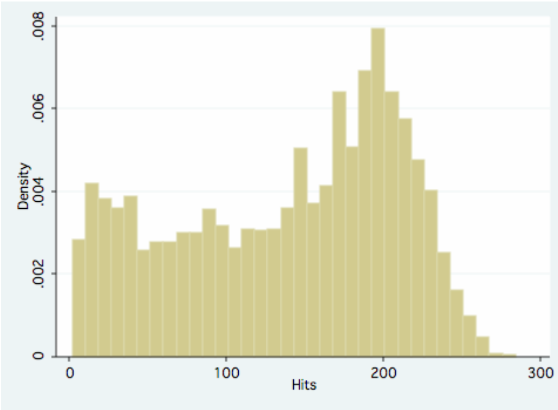
3) Games Started



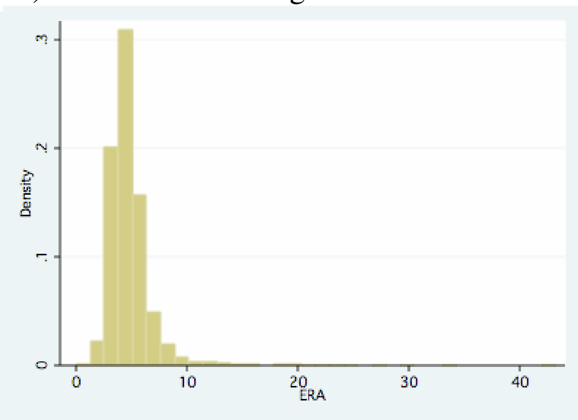
4) Innings Pitched



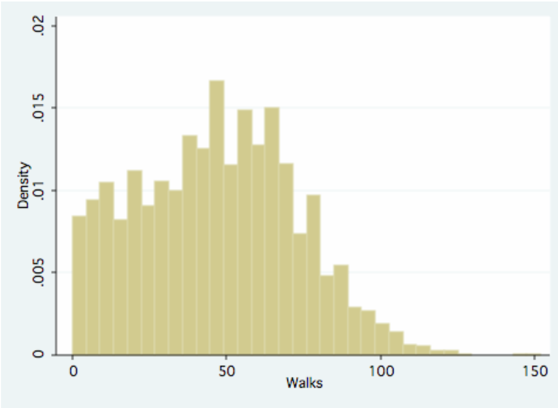
5) Hits



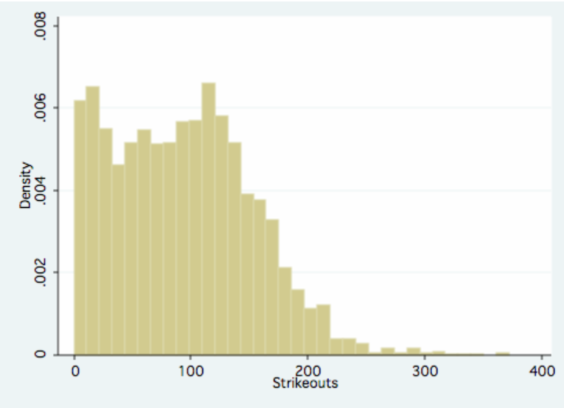
6) Earned Run Average



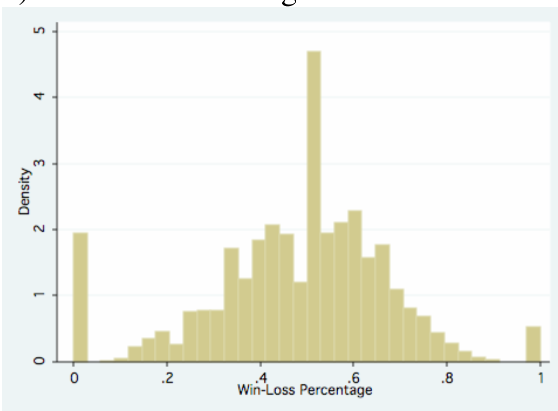
7) Walks



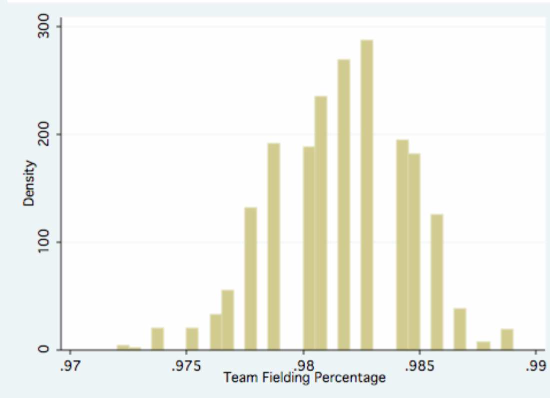
8) Strikeouts



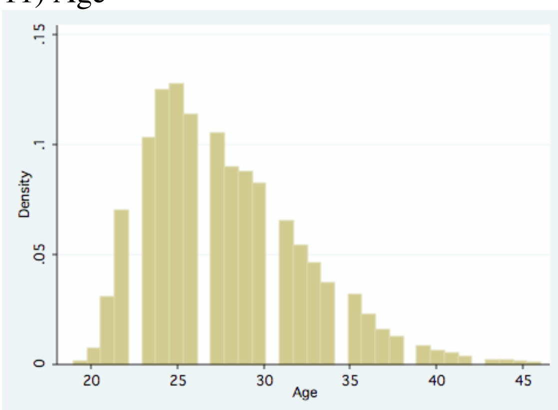
9) Win-Loss Percentage



10) Team Fielding Percentage



11) Age



11) On Base Plus Slugging

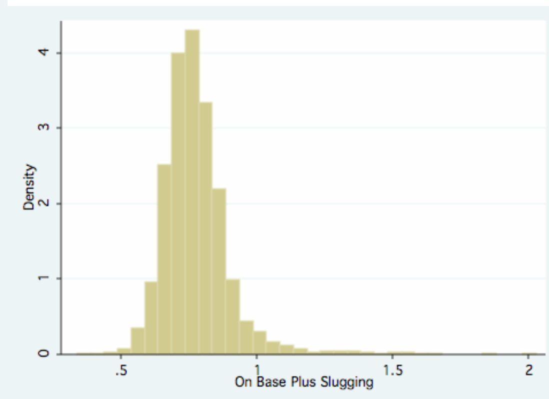


Table II. Uncertainty Model Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Salary	3,533,321	4,115,074	300,000	22,500,000
Team Revenue	156,000,000	40,200,000	80,000,000	327,000,000
Win Percentage	0.499	0.067	0.315	0.648
Innings Pitched	5.859	1.519	0.333	10.000
Hits	6.165	2.270	0.000	15.000
Earned Run Average	5.900	8.382	0.000	270.000
Walks	2.004	1.434	0.000	10.000
Strikeouts	4.067	2.392	0.000	18.000
Team Fielding	0.983	0.002	0.977	0.989
Stadium Type	0.762	0.426	0.000	1.000

Notes: 24,268 Observations

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Feb. 2009.

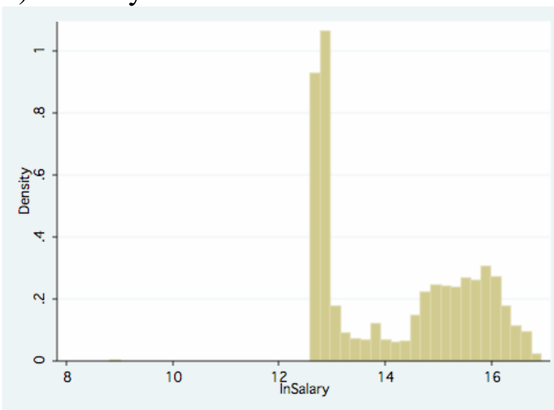
Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Feb. 2009.

Player Index. The Baseball Cube. Major League Baseball. 27 Feb. 2009.

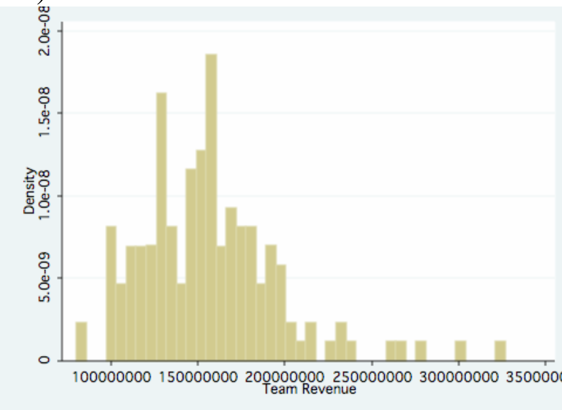
The Business of Baseball. Forbes. Major League Baseball. 27 Feb. 2009.

Figure 2. Histograms

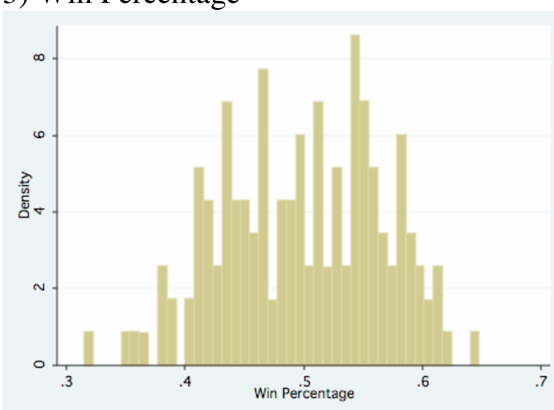
1) lnSalary



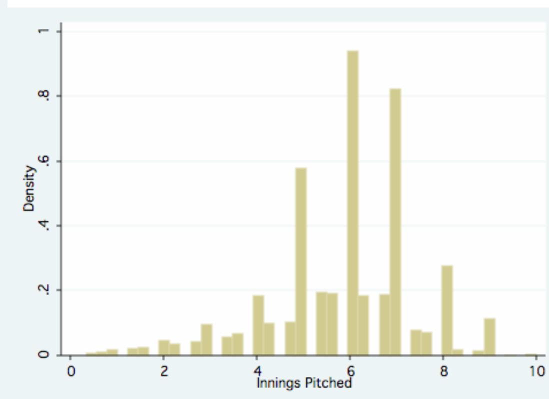
2) Team Revenue



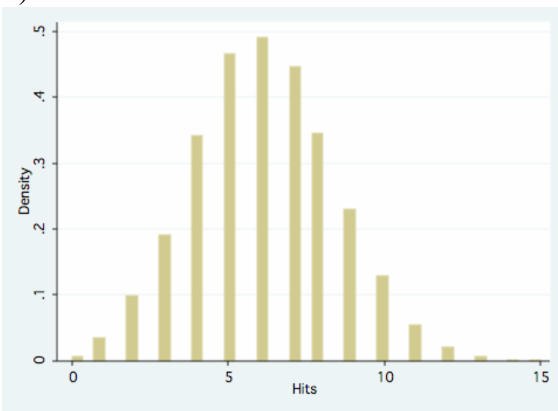
3) Win Percentage



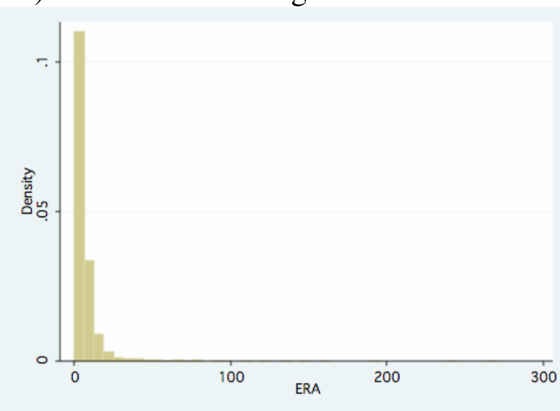
4) Innings Pitched



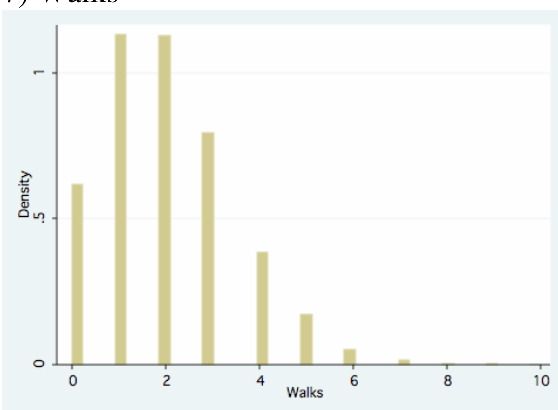
5) Hits



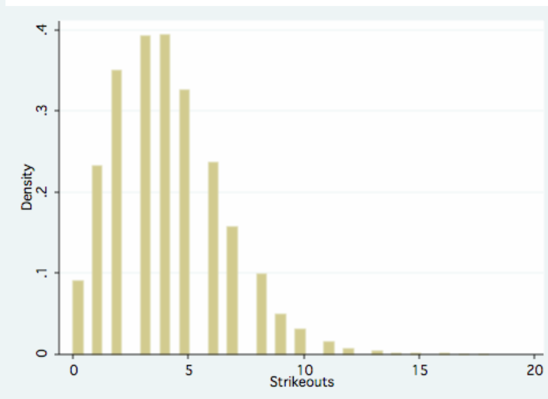
6) Earned Run Average



7) Walks



8) Strikeouts



9) Team Fielding Percentage

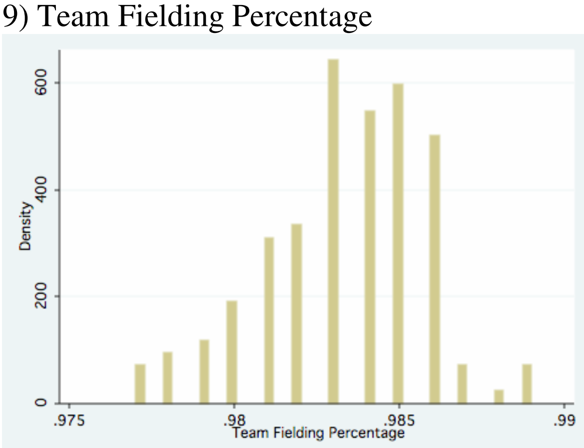


Table III. Multicollinearity

	lnSalary	lnRevenue	Games Started	Innings Pitched/Game	Hits/Game	Earned Run Average	Walks/Game
lnSalary	1						
lnRevenue	0.3688	1					
Games Started	0.411	0.0703	1				
Innings Pitched/Game	0.3615	-0.0174	0.6687	1			
Hits/Game	0.2263	-0.005	0.2937	0.41	1		
Earned Run Average	-0.1709	0.0041	-0.4377	-0.6679	0.1081	1	
Walks/Game	-0.1196	-0.0449	-0.0513	-0.0246	-0.1796	0.0958	1
Strikeouts/Game	0.324	0.1017	0.4768	0.6515	0.0186	-0.4555	0.11
Win-Loss %	0.2095	0.1038	0.3999	0.5153	-0.0606	-0.5668	-0.0701
Team Fielding %	0.2004	0.3979	0.0636	0.0386	-0.0233	-0.0393	-0.0365
Park Factor	-0.0054	-0.0115	-0.021	-0.045	0.0696	0.0495	0.0183
Age	0.6333	0.1245	0.2169	0.1328	0.1616	-0.0347	-0.1773
Age^2	0.6104	0.1287	0.202	0.1231	0.149	-0.0322	-0.1695
On Base Plus Slugging	-0.1851	0.0385	-0.4485	-0.6815	0.2551	0.8588	0.0739
Singed Contract	0.1395	0.0339	-0.0251	-0.0771	0.0283	0.0828	-0.0661

	Strikeouts/Game	Win-Loss %	Team Fielding %	Park Factor	Age	Age^2	On Base Plus Slugging
lnSalary							
lnRevenue							
Games Started							
Innings Pitched/Game							
Hits/Game							
Earned Run Average							
Walks/Game							
Strikeouts/Game	1						
Win-Loss %	0.4382	1					
Team Fielding %	0.0595	0.1498	1				
Park Factor	-0.0684	-0.0342	0.0193	1			
Age	0.0381	0.0598	0.0583	-0.0103	1		
Age^2	0.0365	0.0603	0.0571	-0.0128	0.9947	1	
On Base Plus Slugging	-0.558	-0.5883	-0.0244	0.0964	-0.0344	-0.0325	1
Singed Contract	-0.0941	-0.0742	-0.0001	0.0306	0.3797	0.3721	0.08

Table IV. Perfect Certainty Model Regression Results

	Regression Coefficients			
Variable	(i)	(ii)	(iii)	(iv)
Park Factor	**0.173**	-	**0.485**	**-0.02**
	(1.42)	-	**(2.46)***	**(0.13)**
Games Started	0.012	0.012	0.013	-0.011
	(5.53)**	(5.53)**	(4.18)**	(3.66)**
Innings Pitched	0.14	0.138	0.126	0.169
	(2.74)*	(2.71)*	(1.84)	(2.25)*
Hits	0.076	0.078	0.084	0.065
	(2.05)*	(2.10)*	(1.63)	(1.22)
Earned Run Average	0.05	0.049	0.062	0.023
	(3.26)**	(3.19)**	(3.17)**	(2.74)*
Walks	-0.009	-0.008	-0.019	0.004
	(0.33)	(0.28)	(0.47)	(0.10)
Strikeouts	0.179	0.179	0.188	0.161
	(10.58)**	(10.54)**	(7.64)**	(6.80)**
Win-Loss Percentage	0.013	0.018	0.06	0.093
	(0.12)	(0.17)	(0.39)	(0.64)
Team Fielding Percentage	14.131	14.166	0.689	2.895
	(3.27)**	(3.29)**	(0.10)	(0.47)
lnTeam Revenue	0.754	0.754	0.742	0.802
	(23.98)**	(23.94)**	(16.73)**	(17.92)**
Age	0.771	0.772	0.748	0.781
	(22.67)**	(22.69)**	(14.97)**	(17.14)**
Age^2	-0.009	-0.009	-0.009	-0.009
	(17.50)**	(17.52)**	(11.61)**	(13.09)**
On Base Plus Slugging	-0.545	-0.52	-1.609	-0.337
	(1.36)	(1.30)	(1.04)	(0.62)*
Signed Contract	-0.302	-0.3	-0.221	-0.42
	(5.90)**	(5.86)**	(3.09)**	(5.61)**
Constant	-5.527	-5.381	-5.344	-5.94
	(9.03)**	(8.93)**	(5.92)**	(7.14)*

Notes: 3,083 Observations — Adjusted $R^2 = 0.599$

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level

Newey-West Corrected Standard Errors used for each regression

(i) = Main regression

(ii) = Regression with no Park Factor

(iii) = American League

(iv) = National League

Table V. Perfect Certainty Model AL Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Salary	2,166,278	3,071,211	100,000	19,000,000
Team Revenue	107,158,800	57,791,620	26,000,000	327,000,000
Games Started	21.539	10.812	1.000	36.000
Innings Pitched	133.811	73.185	1.200	271.100
Hits	139.513	72.914	2.000	284.000
Earned Run Average	5.085	2.479	0.000	43.200
Walks	46.454	26.157	0.000	152.000
Strikeouts	88.956	58.158	0.000	313.000
Win-Loss Percentage	0.481	0.195	0.000	1.000
Team Fielding Percentage	0.982	0.003	0.975	0.989
Age	27.861	4.746	19.000	46.000
On Base Plus Slugging	0.787	0.129	0.436	2.029
Signed Contract	0.179	0.383	0.000	1.000
Park Factor	1.011	0.119	0.515	1.584

Notes: 1,516 Observations

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Oct. 2009.

Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Oct. 2009.

Player Index. The Baseball Cube. Major League Baseball. 27 Oct. 2009.

The Business of Baseball. Forbes. Major League Baseball. 27 Oct. 2009.

MLB Franchise Revenues and Expenses. Baseballchronology.com. http://www.baseballchronology.com/Baseball/Years/2005/Income_Expenses.asp. 27 Oct. 2009.

Table VI. Perfect Certainty Model NL Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Salary	2,237,882	3,139,983	100,000	22,000,000
Team Revenue	104,982,200	50,683,530	24,900,000	235,000,000
Games Started	22.162	10.578	1.000	37.000
Innings Pitched	136.988	71.046	0.000	271.200
Hits	137.329	68.402	3.000	268.000
Earned Run Average	4.711	3.168	1.020	99.990
Walks	47.092	25.410	0.000	128.000
Strikeouts	96.923	60.963	0.000	372.000
Win-Loss Percentage	0.478	0.193	0.000	1.000
Team Fielding Percentage	0.981	0.003	0.972	0.989
Age	27.715	4.312	19.000	46.000
On Base Plus Slugging	0.767	0.125	0.340	2333.000
Signed Contract	0.154	0.361	0.000	1.000
Park Factor	0.992	0.140	0.540	1.609

Notes: 1,567 Observations

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Oct. 2009.

Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Oct. 2009.

Player Index. The Baseball Cube. Major League Baseball. 27 Oct. 2009.

The Business of Baseball. Forbes. Major League Baseball. 27 Oct. 2009.

MLB Franchise Revenues and Expenses. Baseballchronology.com. http://www.baseballchronology.com/Baseball/Years/2005/Income_Expenses.asp. 27 Oct. 2009.

Table VII. Multicollinearity

	lnSalary	Team Revenue	Win Percentage	Strikeouts	Walks
lnSalary	1				
Team Revenue	0.2576	1			
Win Percentage	0.2530	0.3614	1		
Strikeouts	0.1061	0.0314	0.0711	1	
Walks	-0.0753	-0.0119	-0.0883	-0.0679	1
Hits	0.0753	-0.0237	-0.0078	-0.1971	-0.1764
Innings Pitched	0.1720	0.0050	0.0909	0.4377	-0.1752
Earned Run Average	-0.0540	-0.0167	-0.0553	-0.3046	0.0588
Team Fielding Percentage	0.1492	0.2407	0.3978	0.0322	-0.0319
Stadium Type	0.0168	0.0757	0.0310	-0.0190	0.0209

	Hits	Innings Pitched	Earned Run Average	Stadium Type	Team Fielding Percentage
lnSalary					
Team Revenue					
Win Percentage					
Strikeouts					
Walks					
Hits	1				
Innings Pitched	-0.1310	1			
Earned Run Average	0.2492	-0.6251	1		
Team Fielding Percentage	-0.0067	0.0591	-0.0354	1	
Stadium Type	0.0112	-0.0269	0.0175	0.0007	1

Table VIII. Uncertainty Model Regression Results

	Regression Coefficients		
Variable	(i)	(ii)	(iii)
Stadium Type	**0.0678**	**-0.0220**	**0.1446**
	(2.93)**	**(0.50)**	**(3.30)****
Team Revenue	0.0042	0.0046	0.0043
	(6.38)**	(12.24)**	(10.76)**
Win Percentage	-0.5683	-0.4379	-0.5431
	(1.32)	(2.13)*	(2.37)*
Strikeouts	0.0324	0.0179	0.0502
	(6.70)*	(2.84)*	(7.20)**
Walks	-0.0240	-0.0171	-0.0310
	(3.52)**	(1.92)	(3.09)**
Hits	0.0554	0.0478	0.0632
	(12.81)**	(8.22)**	(10.30)**
Inning Pitched	0.1768	0.1995	0.1599
	(19.74)**	(17.46)**	(12.62)**
Earned Run Average	0.1050	0.0128	0.0089
	(6.59)**	(6.70)**	(4.03)**
Team Fielding Percent	0.3541	1.5897	6.6788
	(0.03)	(0.32)	(1.20)
Constant	12.6762	12.6889	12.6288
	(175.55)**	(133.93)**	(121.93)**

Notes: 24,263 Observations $R^2 = 0.150$

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level

Newey-West Corrected Standard Errors used for each regression

(i) = Main regression

(ii) = National League

(iii) = American League

Table IX. Uncertainty Model NL Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Salary	3,416,475	3,945,314	300,000	22,500,000
Team Revenue	155,000,000	35,700,000	80,000,000	327,000,000
Win Percentage	0.494	0.063	0.315	0.648
Innings Pitched	5.872	1.486	0.333	10.000
Hits	6.073	2.235	0.000	15.000
Earned Run Average	5.658	7.577	0.000	189.000
Walks	2.055	1.442	0.000	10.000
Strikeouts	4.129	2.409	0.000	18.000
Team Fielding	0.983	0.002	0.977	0.989
Stadium Type	0.803	0.398	0.000	1.000

Notes: 12,943 Observations

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Feb. 2009.

Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Feb. 2009.

Player Index. The Baseball Cube. Major League Baseball. 27 Feb. 2009.

The Business of Baseball. Forbes. Major League Baseball. 27 Feb. 2009.

Table X. Uncertainty Model AL Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Salary	3,664,479	4,297,337	300,000	22,500,000
Team Revenue	158,000,000	44,800,000	80,000,000	327,000,000
Win Percentage	0.505	0.071	0.315	0.648
Innings Pitched	5.845	1.555	0.333	10.000
Hits	6.269	2.305	0.000	15.000
Earned Run Average	6.178	9.208	0.000	270.000
Walks	1.948	1.425	0.000	9.000
Strikeouts	3.995	2.370	0.000	17.000
Team Fielding	0.983	0.002	0.977	0.989
Stadium Type	0.715	0.451	0.000	1.000

Notes: 11,325 Observations

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Feb. 2009.

Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Feb. 2009.

Player Index. The Baseball Cube. Major League Baseball. 27 Feb. 2009.

The Business of Baseball. Forbes. Major League Baseball. 27 Feb. 2009.

Appendix A: Theory

Derivation of basic model:

$V_1 = U(Y_{11}) + \beta U[Y_{12}(x(g))]$
$V_2 = U(Y_{21}) + \beta U[Y_{22}(x(g))]$

Where x = statistics prior to current period, and g = stadium effect

The pitcher will therefore only choose V_1 if:
$V_1 \geq V_2$

Given the assumptions described in the theory section:
$Y_{12} < Y_{22} \Rightarrow \mathbf{Y_{11} > Y_{21}}$

Set $V_1 = V_2$, solve for $U(Y_{11})$

$U(Y_{11}) = U(Y_{21}) + \beta[U(Y_{22}) - U(Y_{12})]$

To solve for the effect of the hitter's ballpark on salary, I assume the following functional form for utility:
$U(Y_{ij}) = Y_{ij}$

The following steps are therefore used to derive the model:
$Y_{11} = Y_{21} + \beta(Y_{22}(x(g)) - Y_{12}(x(g)))$

From here we may take the derivative and solve for $\partial Y_{11}/ \partial g$:
$\partial Y_{11}/ \partial g = \beta[\partial Y_{22}/ \partial x * \partial x/ \partial g - \partial Y_{12}/ \partial x * \partial x/ \partial g$

In the neutral stadium, the stadium effect (g in Y_{22}) is equal to zero and it is also considered zero in the hitter's stadium. Therefore, the final equation is the following:
$\boldsymbol{\partial Y_{11}/ \partial g = 0}$

Derivation of model with uncertainty:

$V_1 = U(Y_{11}) + \beta EU(Y_{12})$
$V_2 = U(Y_{21}) + \beta U(Y_{22})$

Where E = expected value and $Y_{12} = [p(g)(Y_{12}^{Low}(x(g))) + (1-p(g))(Y_{12}^{High}(x(g)))]$

The pitcher will therefore only choose V_1 if:
$V_1 \geq V_2$

Given the assumptions described in the theory section:
$Y_{12} < Y_{22} \Rightarrow \mathbf{Y_{11} > Y_{21}}$

Set $V_1 = V_2$, solve for $U(Y_{11})$

$U(Y_{11}) = U(Y_{21}) + \beta[U(Y_{22}) - E(U(Y_{12}))]$

To solve for the effect of the hitter's ballpark on salary, I assume the following functional form for utility:
$U(Y_{ij}) = Y_{ij}$
$Y_{11} = Y_{21} + \beta(Y_{22} - E(Y_{12}))$

From here we may take the derivative and solve for $\partial Y_{11}/ \partial g$:
$\partial Y_{11}/ \partial g = \beta[\partial Y_{22}/ \partial x * \partial x/ \partial g - ((pr(g)(\partial Y_{12}^{L}/\partial x * \partial x/ \partial g * g) - (1 - pr(g))(\partial Y_{12}^{H}/\partial x * \partial x/ \partial g * g)))]$

In the neutral stadium, the stadium effect (g in Y_{22}) is equal to zero. Therefore, the final equation is the following:
$\boldsymbol{\partial Y_{11}/ \partial g = -\beta[((p(g)(\partial Y_{12}^{L}/\partial x * \partial x/ \partial g *) + (1 - p(g))(\partial Y_{12}^{H}/\partial x * \partial x/ \partial g * g)))]}$

and where $\partial x/ \partial g < 0$.

Appendix B: Perfect Certainty Estimation Issues

I tested the data for various estimation issues. The data are organized as unbalanced panel data, where there is a panel for each season of a pitcher's career. There is multicollinearity seen in this data set. The multicollinearity however, did not lead me to remove any variables. Each variable has a different effect with regard to the pitcher's value. To address this, I transform the multicollinear variables (strikeouts, walks, innings pitched, and hits) by dividing them by the total number of games pitched for each pitcher. As can be seen in table III, the new correlation coefficients show that there is little multicollinearity between the variables. One might expect the pitcher ERA and the park factor variables to be highly correlated. They are not because the park factor is only for the pitcher's home games, therefore all away games are reflected in the ERA but not the park factor.

I use the Newey West Standard Errors to correct for the strong serial correlation exhibited by the data. The Breusch-Pagan test finds no heteroskedasticity. The team-fielding variable is the only variable found to contain non-stationarity, and I correct for the non-stationarity by using the first difference. After first differencing, the Durbin Watson test once again indicated serial correlation, which I corrected for by using Newey West Standard Errors.

Appendix C: Uncertainty Model Estimation Issues

The data are again organized as unbalanced panel data, where there is a panel for each game of a pitcher's career. As can be seen in table VII, the simple correlations do not indicate multicollinearity. The correlation between innings pitched and ERA returns the most significant coefficient of -.63, with only one other combination of variables reaching 0.43. ERA is partially calculated using innings pitched, which explains the correlation between the two. We consider both necessary in the equation however, and do not eliminate them.

The Durbin-Watson test repeated for each panel, as well as the partial autocorrelation graphs, indicate strong serial correlation, which I corrected by using Newey West standard errors.

On randomly chosen cross sectional units, the Breusch-Pagan test failed to reject the null hypothesis that there is constant variance. Therefore, it can be inferred that there is no heteroskedasticity in the data, and no correction is needed.

I tested for stationarity using the Dickey Fuller test, panel by panel, for every variable included in the regression. The performance variables; strikeouts, walks, hits, innings pitched, and ERA, all returned results that reject the null hypothesis, which states that the variable has a unit root, for most of the panels. The stadium type dummy variable also rejected the null hypothesis for almost all of the panels. The remaining variables, nominal team revenue, team winning percentage, and team fielding percentage, all failed to reject the null hypothesis, therefore leading to the conclusion that they are non-stationary. We also tested the residual for stationarity to determine if there was cointegration, but the residual failed to reject the null hypothesis. This outcome requires the use of first differences for those variables that were found to be nonstationary. I reran the regression using the first differences variables, and retested for serial correlation. The Durbin Watson test and partial autocorrelation graphs once again indicated serial correlation, which I corrected by using Newey West standard errors.

Works Cited

Feinberg, Robert M. "Earnings-Risk as a Compensating Differential." Southern Economic Journal 48.1 (1981): 156-63.

Goodman, G. H., and F. T. McAndrew. "Domes and Astroturf: A Note on the Relationship between the Physical Environment and the Performance of Major League Baseball Players." Environment and Behavior 25.1 (1993): 121.

Kahn, L. M. "Free Agency, Long-Term Contracts and Compensation in Major League Baseball: Estimates from Panel Data." Review of Economics and Statistics 75.1 (1993): 157.

Krautmann, Anthony C., Elizabeth Gustafson, and Lawrence Hadley. "A Note on the Structural Stability of Salary Equations: Major League Baseball Pitchers." Journal of Sports Economics 4.1 (2003): 56-63.

Krautmann, Anthony C., and Paul L. Novak. "Compensating Wage Differentials in Major League Baseball." Indian Journal of Economics and Business 3.2 (2004): 259-68.

McGoldrick, Kimmarie, and John Robst. "The Effect of Worker Mobility on Compensating Wages for Earnings Risk." Applied Economics 28.2 (1996): 221-32.

Moore, Michael J. "Unions, Employment Risks, and Market Provision of Employment Risk Differentials." Journal of Risk and Uncertainty 10.1 (1995): 57-70.

Rosen, Sherwin. "The Theory of Equalizing Differences." Handbook of Labor Economics Volume 1. (1986): 641-692

Rottenberg, Simon. "The Baseball Players' Labor Market." The Journal of Political Economy 64.3 (1956): 242-58.

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

Data Sources

"Major League Baseball Player Data." ESPN. Major League Baseball. <http://sports.espn.go.com/mlb/players>. 27 Oct. 2009.

MLB Franchise Revenues and Expenses. Baseballchronology.com. http://www.baseballchronology.com/Baseball/Years/2005/Income_Expenses.asp. 27 Oct. 2009.

"Player Index." The Baseball Cube. Major League Baseball. 27 Feb. 2009.

Sports Reference LLC. "Game Finder." Baseball-Reference.com - Major League Statistics and Information. http://www.baseball-reference.com/. 27 Oct. 2009.

The Business of Baseball. Forbes. Major League Baseball. 27 Oct. 2009.