

The Relationship between Employee Turnover and Firm Performance: An Analysis of Major League Baseball from 2002-2016

By Benjamin Ruskin
Advised by Professor Richard Ball

Senior Thesis in Economics
Haverford College
April 27th, 2017

ABSTRACT

Does turnover improve performance by allowing firms and employees to optimally match, as outlined by job matching theory? On the other hand, could turnover harm productivity by disrupting team dynamics, as outlined by the Firm Specific Human Capital Model (FSHCM)? I attempt to answer these questions through an analysis of Major League Baseball. For exploring the general relationship between turnover and performance, I regress team turnover rates against their winning percentage using both OLS and quadratic models. For specific theories, I analyze whether positional turnover, inter-league turnover, or the interaction between turnover and ballpark characteristics affect team performance using OLS regression. I attempt to pinpoint precisely *how* job matching theory and FSHCM could be operating in baseball by analyzing these secondary explanatory variables. I find no evidence to suggest that turnover has a significant effect on team performance over a full season. Rather, roster quality and past winning percentage appear to be better indicators of future winning percentage. However, when looking at the effect of turnover over only half the season, it appears that the best teams from the previous season benefit and the worst teams from the previous season are harmed. I attribute this difference to the ability of better teams to attract better players during the off-season.

TABLE OF CONTENTS

I. INTRODUCTION.....	4
II. REVIEW OF RESEARCH ON EMPLOYEE TURNOVER	7
III. APPLYING TURNOVER THEORY TO BASEBALL	13
IV. BACKGROUND ON BASEBALL PROJECTION SYSTEMS	17
V. DATA.....	19
VI. METHODOLOGY	
A) Model of General Turnover Effects	25
B) Secondary Models of Turnover	29
C) Alternative Measures of Key Variables	31
VII. RESULTS	
A) Model of General Turnover Effects	33
B) Secondary Models of Turnover	39
C) Alternative Measures of Key Variables	42
VIII. DISCUSSION.....	48
IX. CONCLUSION.....	52
X APPENDICES	
A) Data Appendix.....	54
B) Robustness Appendix.....	75
XI. REFERENCES	80
XII. DATA SOURCES	82

I. INTRODUCTION

Despite a significant amount of literature on the effects of employee turnover rates on wages, unemployment and more, there has been very little research on the effect of turnover on the performance of the firm itself. I analyze this relationship through the prism of Major League Baseball. Surprisingly, the research that has been done on this relationship has not been extended to the realm of sports, where team chemistry and individual performance can have a profound impact on the overall performance of the team. Baseball is the ideal sport from which to study employee turnover because each team consists of a roster in which many players are utilized, individual performance is easily quantified, and player communication is vital. Examining whether employee turnover affects firm performance is the central goal of this research.

Two economic theories drive my analysis. The Firm Specific Human Capital theory (FSHCM), proposed by Becker (1975), predicts a negative relationship between turnover and firm performance. Job matching theory, proposed by Jovanovic (1979), predicts a positive relationship between turnover and performance. First, I examine the general relationship between team performance and turnover rates. Even though one may conclude that only either the FSHCM or job matching theory can be correct, there is some evidence that job matching theory may dominate at lower levels of turnover and FSHCM may dominate at higher levels of turnover in other industries (Harris, Tang, and Tseng, 2006; Siebert and Zubanov, 2009). I assert that, given the nature of team dynamics in baseball, this prediction is plausible because low turnover may improve team quality without affecting team chemistry, but high turnover may affect team chemistry to the point that it undermines any improvement.

Furthermore, I extend my analysis beyond previous research by attempting to pinpoint exactly where these two contrasting theories manifest themselves in baseball. I argue that the effects of FSHCM may be seen through positional turnover and the effects of job matching theory can be observed through inter-league turnover. I hypothesize that turnover at shortstop, catcher, and pitcher — positions that involve more signal calling — negatively affect performance because of the extra training involved in teaching replacement players to communicate with their teammates. I also predict that teams that acquire more players from the opposite league will perform better because they are identifying players whose productivities will improve with a league switch due to differences in American League and National League rules. Finally, I predict that, for teams with ballparks whose dimensions and characteristics favor either hitters or pitchers to a significant degree, the effect of turnover on performance is ambiguous, but could be different than teams with standard ballpark characteristics. The effects of FSHCM could be amplified because players need time to get used to playing in a new, unusual ballpark. The effects of job matching theory could be amplified because those teams are more easily able to find players who are optimally matched to their ballpark.

To test these hypotheses, I collected balanced panel data for every Major League Baseball team for every year from 2002 to 2016. I recorded each team's record and winning percentage for each season and its turnover rate from season to season. I also collected data on positional turnover, inter-league turnover, and ballpark effects for each team. I then regressed winning percentage against the general turnover rate, using positional turnover, inter-league turnover, and park effects as secondary explanatory variables. This analysis is conducted using

OLS regression to test for a linear relationship between turnover and performance. I also run a quadratic model for general turnover to test for the predicted inverted U-shaped relationship that implies an interior optimal roster turnover rate. I control for team payroll, managerial experience, average player quality on the opening day roster (measured through wins above replacement, a statistic for player quality), the number of significant players on the roster, and winning percentage in the previous season. I run models both with and without team fixed effects and with and without an interaction between turnover and past performance. I use alternative measures of turnover and team performance to test the robustness of these models, and look at the effect of turnover over both the short-run (half-season) and the medium-run (full-season).

General managers and other baseball executives can use the results of this study in their decision making process regarding player transactions as they attempt to maximize their teams' performance. Walking the fine line between allowing players to develop together as a team and also improving roster quality is a dilemma every general manager faces. Finally, my research adds to the limited academic literature that analyzes the effects of turnover on firm performance through the application FSHCM and job matching theory.

II. REVIEW OF RESEARCH ON EMPLOYEE TURNOVER

As Glebbeek and Bax (2004) point out, when it comes to academic research, there has been a significant amount of time spent analyzing the mechanisms behind employee turnover, but there has been little attention paid to the effect of turnover on firm performance. Research on the latter relationship has centered around two contrasting theoretical models regarding turnover and performance: The Firm Specific Human Capital Model (FSHCM), proposed by Becker (1975), and job matching theory, proposed by Jovanovic (1979). The main conclusion of both theories is that wages and employee tenure have a positive relationship. However, the main difference between the two theories is in the direction of causation (Glenn et al, 2001).

The Firm Specific Human Capital Model (FSHCM) predicts that wages will increase with tenure due to the human capital individuals accumulate with their employers over time through experience, training, and familiarity with the job. Therefore, any employee's productivity at their firm will be higher relative to other potential workers who do not have experience at the firm. Becker equates investment in a person's training and education to business investments. His model predicts that firms bearing the cost of employee training will be less productive if employee turnover rates are high because there will be a lower incentive to properly train new employees. This is a seminal work for my paper because it is the first model of the effects of employee turnover.

Job matching theory predicts that higher wages lead to longer tenure. An employee who is optimally matched with a firm will be more productive for that firm than any other potential worker. Furthermore, the employee will be more productive with that firm than with any other

firm. Over time, that employee's wage will increase as compensation for that productivity. Since the worker's productivity would be lower elsewhere, that wage will not be matched by other firms. Therefore, the employee will remain with the firm. As a result, according to job matching theory, any turnover that does occur would be the result of inefficient matching that does not optimize either the firm's or the worker's productivity. In other words, firms and employees only separate if there is a better match available. Contradicting Becker, this conclusion implies that higher employee turnover rates will increase firm productivity because firms will continue to look for employees until labor productivity is maximized.

My hypothesis that there is an optimal turnover rate that baseball teams should target to maximize performance applies these two theories in tandem. Initially, at lower turnover rates, the marginal benefit of additional turnover from job matching will outweigh the marginal cost described in FSHCM. In baseball, job matching benefits could be a product of finding players specialized to the unique characteristics of that organization. FSHCM costs could include the extra training time necessary in practice or the disruption of team chemistry. As turnover rates increase, increasing marginal cost will eventually overtake the diminishing returns of job matching. The point where these marginal costs and marginal benefits are equal would be the optimal rate of turnover. From there, if turnover rates increased further, marginal cost would outweigh marginal benefit, leading to a negative relationship between turnover and performance.

There has been some previous research demonstrating the validity of this hypothesis in other industries. Harris, Tang, and Tseng (2006) use longitudinal survey data from Australian firms to find a relationship between employee turnover rates and labor productivity. In fact,

they do find that job matching theory dominates when turnover is low and the FSHCM dominates when turnover is high, leading to the inverted U-shape relationship predicted in my hypothesis. They conclude that there is an optimal turnover rate of 22 percent at which these firms can maximize performance. Siebert and Zubanov (2009) extend this finding when they study the relationship between employee turnover and labor productivity using five years of panel data from stores belonging to a single U.K. clothing company. Unlike previous research, they are able to separate out part-time workers from full-time workers. They find that part-time worker turnover and productivity display the inverted U-shape relationship found in Harris, Tang, and Tseng (2006), and that full-time worker turnover has a negative relationship with firm productivity. This finding supports the intuition that continuity amongst full-time workers is more important than amongst part-time workers. I propose that, given the high levels of player mobility in Major League Baseball, with players being traded or released without any notice, they will exhibit the characteristics of part-time workers.

Even though there is some evidence of an inverted U-shaped relationship where job matching theory dominates when turnover rates are low and FSHCM dominates when turnover rates are high, there is also ambiguous evidence from the single company study conducted by Glebbeek and Bax (2004). They compared different branches from within the same organization and were unable to clearly establish an optimal point of turnover. However, they do determine that their findings refute the possibility of a negative relationship between the two variables, contradicting the findings of Seibert and Zubanov (2009). They conclude that using different branches of the same organization allows for the greatest amount of control in analyzing whether there is an optimal rate of turnover.

Therefore, the sports industry provides an effective medium for examining the relationship between turnover and performance, given the ease with which one can look at data from different teams within one sports league. However, researchers have not taken advantage of this opportunity. Roster turnover has never been used as an explanatory variable in predicting team performance.

The relationship between managerial turnover and team performance has been examined in Major League Baseball (Hill, 2009). In this study, it is concluded that team performance is negatively affected by managerial turnover in baseball. Small amounts of change may have a short-term positive effect, but the law of diminishing returns is illustrated as managerial change becomes more frequent within an organization (Hill, 2009). This study illustrates the importance of controlling for managerial turnover when looking at roster turnover in sports.

Furthermore, the relationship between roster turnover and team attendance has been examined using Major League Baseball data (Kahane and Shmanske 1997), and National Basketball Association data (Morse, et al, 2008). Attendance in Major League Baseball is negatively associated with roster turnover, but there is no association in the National Basketball Association (Morse, et al, 2008). The reason for the difference in these findings may be attributed to the characteristics of the two sports. Baseball teams cannot rely on star players to the degree that basketball teams do because only one player bats at a time. In basketball, a star player can play for the entire game. As a result, average turnover in basketball may be a less reliable predictor of team performance than average turnover in baseball, just as it had less of an impact on attendance. The success of teams in basketball is generally dependent on star

players. For example, when Basketball MVP LeBron James left the Cleveland Cavaliers, the team's winning percentage fell from .744 to .232. When he returned, it improved from .402 to .646 (Land of Basketball, n.d.). In basketball, the turnover of a single player can lead to dramatic fluctuations in performance because a team can rely on that star player to carry the team. Therefore, it is plausible that average turnover in basketball and sports like football (where the team relies heavily on the quarterback) may be far less important than the type of turnover that occurs. Because the game of baseball involves each individual player performing actions independently (players take turns hitting and pitching), the loss of a single player will not have the same effect on overall performance. Given that this study is interested in the relationship between the general turnover rate and performance, baseball appears to be the most appropriate sport through which to conduct my analysis.

There is additional evidence in the literature to suggest that baseball is an appropriate medium for analyzing the relationship between employee turnover and firm performance. Glebbeek and Bax (2004) point out that one of the difficulties in testing this relationship is controlling for changes in employee quality as a result of turnover. In exploring the relationship between turnover and attendance, Kahane and Shmanske (1997) attempted to control for quality using changes in salary. However, they admitted that, because younger, high quality players are not paid nearly as well as older players of all qualities, this method for controlling for changes in team quality is unreliable. They hoped that a better alternative could be used in future analysis of turnover. Since their research was published, Wins Above Replacement (WAR), a measure of player quality used frequently in baseball, has been introduced. Using this new statistic is likely the type of extension the researchers envisioned, given its accuracy in

converting player level data into team wins. WAR compares a player's performance, through runs contributed, to league averages and replacement level players at his position. As a result, now one can more accurately isolate players of similar quality and analyze the effect of their turnover on performance. In addition, there are no statistics used in other industries that convert employee quality to firm profits so cleanly, further illustrating the rationale for using sports as a medium for answering my research question.

III. APPLYING TURNOVER THEORY TO BASEBALL

Not only is there limited research on the relationship between general turnover rates and firm productivity in other industries, but no study has aimed to pinpoint potential sources of the effect turnover. Therefore, my analysis extends beyond previous work in two different ways. First, it is able to analyze the turnover of many firms within one industry by taking advantage of the availability of sports data. Secondly, because of this abundance of data, I can attempt to pinpoint exactly *where* FSHCM and job matching theory manifest themselves within the turnover story.

In order to determine the specific areas of baseball in which to examine in this approach, it is important to precisely differentiate between the two theories. FSHCM applies to productivity increases that are the result of the accumulation of capital by players who stay with a team for an extended period of time. On the other hand, job matching theory applies to productivity increases that could only have occurred by matching specific players with teams with certain characteristics that allow them to perform optimally. Therefore, one would expect to see job matching theory apply to observable, constant differences at the team level and FSHCM to apply to observable, constant differences at the player level.

For example, at the player level, certain positions require more communication, which improves over time. Therefore, one would also expect to see FSHCM operate at those positions. In fact, it has been hypothesized that players in positions that require the most team specific knowledge are the least frequently traded because of the high human capital they have accumulated over time with that individual team compared to an equivalent player on the market (Glenn, McGarrity, and Weller, 2001). In baseball, human capital is most significant at

positions that are required to signal instructions to other members of the team on defense. For example, catchers relay pitch selections from the manager and even make decisions themselves on what pitch the pitcher should throw. Their communication with the pitcher is paramount to the success of the team. Based on the pitch selection, the shortstop will often signal to the rest of the team what type of defensive alignment it should take. Therefore, one would expect turnover amongst catchers, pitchers, and shortstops to have a negative impact on team performance, according to FSHCM. As a result, it is no surprise that, from 1900-1992, catchers and shortstops were traded significantly less often than other positions (Glenn, McGarrity, and Weller, 2001). They are traded about 12.5 percent of the time, which is a statistically significant difference from the 16 percent clip at which other positions are traded. The study did not include pitcher turnover in its analysis. It appears as if teams apply the FSHCM to their turnover decisions, especially as it relates to catchers and shortstops.

On the other hand, job matching occurs in situations where certain employers are better suited to certain employees than others. Therefore, the implicit assumption necessary for job matching to take place is that employers are distinct from one another and employees will have different productivity levels depending on the organization for which they work. Does baseball fit this assumption? Compared to other industries, baseball organizations are remarkably uniform. However, there are two main characteristics that differentiate baseball teams: league and ballpark.

Each team either plays in the National League or the American League. For National League teams, the majority of games involve pitchers also hitting. For American League teams, the majority of games involve a designated hitter (a player who only hits and does not play the

field), replacing the pitcher in the lineup. Pitchers generally perform better in the National League because they have to face one less competent hitter (on average, pitchers are extremely poor hitters). Therefore, National League teams may be able to find pitchers currently playing in the American League who could improve their productivities in the National League. On the other hand, some older hitters may be better suited to the American League because they can rest when their team is in the field and focus only on hitting. These players can improve their productivities by switching leagues. Based on job matching theory, one would expect to see teams benefit from bringing in more players from the opposite league, as their productivities may increase relative to the previous season from which their salary is generally based.

It is also possible that either job matching theory or FSHCM may also apply to teams that play their home games in unusual ballparks. Each team plays half of their games in a stadium unique in both dimensions and elevation. Some stadiums have quirks that make them more suitable to a specific type of player, or that require specific training to optimize performance. For example, the Colorado Rockies' stadium, Coors Field, is played at almost one mile above sea level. Because the air is so thin, fly balls carry farther and the stadium yields more home runs than any other park. Therefore, it is plausible that fly-ball hitters will generally improve their productivities with the Rockies and the performances of fly-ball pitchers will regress. In this case, turnover would help the Rockies through job matching. However, it is also plausible that players learn to hit more fly balls, and pitchers learn to force hitters into ground balls more often over time. In this case, turnover would hurt the Rockies through FSHCM.

Therefore, even though it seems reasonable to hypothesize that ballparks have an effect on turnover, it is difficult to predict the direction of the effect.

Luckily, baseball statisticians have developed a quantitative measure of ballpark quiriness called “park factor”. This statistic is a measure of how much offense was produced in the ballpark compared to what one would expect to see in an average ballpark. Teams with high park factors play in ballparks that produce more offense than average and teams with low park factors play in ballparks that produce less offense than average (Fangraphs.com). Sure enough, Coors field has, by far, the highest park factor of any ballpark, confirming what the eye-test suggested before the measure was developed. By using the park factors statistic, I can measure whether teams on the extreme ends of the data benefit more from turnover than teams with park factors closer to the average. Such a result would confirm the presence of job matching theory in baseball. The opposite would confirm the presence of FSHCM.

My research is both a synthesis of the indirectly related studies on the topic of either sports or turnover and an extension of previous research on turnover and performance in that I use an extremely reliable measure of performance, sports wins, and a well-accepted measure of employee quality (WAR) to conduct my analysis. Using this technique allows me to control for player quality. Therefore, I am able to effectively determine the general relationship between turnover and performance. Furthermore, because of the measurability of firm and player differences in baseball, I am able to more precisely identify how Jovanovic’s job matching theory and Becker’s FSHCM map the observed effects of employee turnover.

IV. BACKGROUND ON BASEBALL PROJECTION SYSTEMS

Even though this research is the first of its kind to study the effects of preseason player turnover on in-season team performance in Major League Baseball, it is far from the first to propose a model through which one can estimate a team's win percentage. Fangraphs.com, a baseball statistics database and informational website, lists eleven mainstream projection systems. Generally speaking, these models estimate individual player performance based on past performance, age, and historical trends. How exactly these different factors are weighted is what differentiates the various projection systems (Major League Baseball, n.d.). Some of these projection systems have been used to extrapolate team performance for sites available through subscription such as Baseball Prospectus and FanGraphs (Druschel, 2016). These models, the most prominent baseball projection systems, are the Player Empirical Comparison and Optimization Test Algorithm (PECOTA) and the Szymborski Projection System (ZiPS). Both systems use past performance and age to identify similar players and project their performance through this comparison; however, ZiPS relies more heavily on analytics used to determine how likely it is for balls put in play to turn into hits (Druschel, 2016). This differentiation provides a clear example of how subtle the differences between projection systems are and how complex they appear to be. However, the exact methodologies behind ZiPS and PECOTA are not available to the public, as they are used for profit by their parent companies. Simpler projection systems whose methodologies are available, such as Marcel, have not been taken from the player level to project team performance. Overall, these complex projection systems do not provide an avenue through which to conduct my own analysis based on both their extreme complexity and lack of transparency.

However, Bill James, a baseball statistician, developed a far simpler technique for projecting team performance. Named the Pythagorean Expectation, it is a simple formula that converts runs scored and runs allowed by a team in the past into an expected future winning percentage. The original formula is as follows:

$$\text{Winning Percentage} = \frac{(\text{Runs Scored})^2}{(\text{Runs Scored})^2 + (\text{Runs Allowed})^2}$$

It has generally been used to predict a team's winning percentage for the rest of a season based on performance earlier in that season, as opposed to projecting a team's winning percentage across seasons (Sports Reference LLC, n.d.). Furthermore, the formula itself has been refined over the years to:

$$\text{Winning Percentage} = \frac{(\text{Runs Scored})^{1.83}}{(\text{Runs Scored})^{1.83} + (\text{Runs Allowed})^{1.83}}$$

This refinement has been shown to improve the formula's accuracy (Davenport and Woolner, 1999). The theory behind the Pythagorean Expectation Model is that run differential provides a more accurate representation of a team's performance than actual winning percentage (Sports Reference LLC, n.d.). As a result, in the interest of robustness, it can be used as an alternative measure for past team performance in my own model.

V. DATA

The final dataset used to conduct this analysis is comprised of balanced panel data consisting of one observation for each of the 30 Major League Baseball teams for every year from 2002 to 2016, resulting in a total of 450 observations. Data for 20 other variables were recorded.

“turnover” represents the proportion of significant players from the previous season who are not on the Opening Day roster. Significant players are defined as hitters who appear in at least 100 games and pitchers who either started at least 15 games or appeared in 40 games. The concept of a significant player is taken from Kahane and Shmanske (1997). When analyzing the relationship between turnover and attendance in Major League Baseball, they only looked at the turnover of players who matched criteria similar to those used in my research. The purpose of looking only at significant players is to ensure that my measure of turnover captures the changing of players who have had a significant impact on team performance throughout the previous season. Kahane and Shmanske defined significant players as hitters who appeared in at least 100 games and pitchers who appeared in at least 30 games. I chose to modify the criteria for being a significant pitcher in my study because pitchers fall into two categories — starters and relievers —and applying the same appearance cut-off to these two vastly different groups did not seem appropriate. Influential relievers will generally appear in 60-80 games throughout a season if they avoid injury. Influential starters will generally start 30 games if they avoid injury. By differentiating between starters and relievers, and lowering the threshold to 15 games started and 40 appearances respectively, I allow for pitchers, who are generally more injury prone than hitters, to miss a portion of the regular season and still be deemed significant

if they appear in about half of the games they would have played in had they been healthy for the full season. The turnover rate of significant players on a roster functions as the main explanatory variable in this research.

“rostercont” can replace turnover as the main explanatory variable in the analysis, as it essentially functions as an inverse of total game-level roster turnover without eliminating non-significant players. The variable is defined as the the proportion of games played from the previous season retained by a team in the next season. In other words, on Opening Day, if a team retains players who appeared in 800 of the 1000 appearances made by their entire roster in the previous season, *rostercont* = .80 or 80 percent.

The dependent variable used in this analysis is a team’s winning percentage, or *“WL”*. *“W”* represents the number of games won by each team in a season. *“L”* represents the number of games lost by each team in a season. Therefore:

$$WL = \frac{W}{W + L}$$

“pythWL” is an alternative measure of productivity calculated using the formula:

$$pythWL = \frac{(Runs\ Scored)^{1.83}}{(Runs\ Scored)^{1.83} + (Runs\ Allowed)^{1.83}}$$

where *“Runs Scored”*= the total number of runs scored by the team in a season and *“Runs Allowed”*= the total number of runs the opposition scored against the team in a season.

“AllStarWL” is a measure of short term performance up until the All Star Break. It is the winning percentage of each team up until that point in the season. Usually, the All Star Game is held during the second full week of July. At that point, teams have played anywhere from 80 to 90 games in the regular season, or just over half of their schedule.

“man_tenure” in year t represents the number of full seasons the team’s manager had managed the team prior to t .

“OpenWAR” represents the sum of the WAR for each significant player on every team’s Opening Day roster for every season and *“warexists”* represents the total number of significant players on a team’s Opening Day roster. Using these two variables *“warpersig”*, or the average WAR associated with a significant player on a team’s Opening Day roster, can be calculated using the formula:

$$warpersig = \frac{OpenWAR}{warexists}$$

“vegasW” is an alternative measure of projected team quality based on the over/under betting win total determined by Las Vegas odds-makers. The over/under projected win total reflects how many games they expect each team to win in a given season. ItThis data was collected from www.sportsoddshistory.com. It is important to note that there is likely some variation between the data collected here and the win totals given by individual casinos. Only data from 2004-2016 were available, meaning that I lose one season of observations when I use *“vegasW”* in my analysis.

“totalplayers” represents the total number of players on a team’s Opening Day roster. Using *totalplayers* and *warexists*, the variable *“sigrate”*, or the proportion of significant players on a team’s Opening Day Roster, can be calculated using the formula:

$$sigrate = \frac{warexists}{totalplayers}$$

“teampayroll” represents the estimated total amount of money each team paid their players based on their Opening Day rosters, with the units in millions of dollars.

The following five variables function as secondary explanatory variables to turnover:

“catchercont” represents the proportion of catching appearances from the previous season retained by the team during the next season.

“sscont” represents the proportion of shortstop appearances from the previous season retained by the team during the next season.

“startpitchcont” represents the proportion of pitching starts by the 5 most frequent starters from the previous season retained by the team.

“leagueturnover” represents the number of significant players on a team’s Opening Day Roster who played for a team from the opposite league in the previous season.

“parkfactor” represents the measure of the quirkiness of a team’s ballpark in terms of batter or pitcher friendliness. Each team’s ballpark is assigned a pitcher park factor, or *“ppfactors”*, representing the stadium’s pitcher friendliness. It is also assigned a batter park factor, or *“bpfactors”*, representing the stadium’s hitter friendliness. A neutral ballpark is deemed to have a *ppfactors* and *bpfactors*=100. Therefore, I calculate *“parkfactor”*, or the quirkiness of a team’s ballpark, in relation to this generic standard using the formula:

$$parkfactor = \left| \frac{ppfactors + bpfactors}{2} - 100 \right|$$

Unless otherwise noted, all data was collected from *baseball-reference.com* except for the Opening Day rosters, which were collected from *USA Today MLB Salary Database*.

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
turnover	420	.3165031	.1156928	.047619	.8333333
rostercont	420	.6473264	.1028704	.2125435	.877885
catchcont	420	.6817092	.2823119	0	1
sscont	420	.7111246	.326939	0	1
startpitchcont	420	.7217527	.1916364	0	1
warexists	420	16.96429	2.955705	4	25
sigrate	420	.6146677	.1116753	.1538462	.8928571
warpersig	420	1.875837	.5484971	.1615385	3.65
leagueturnover	420	1.895238	1.458569	0	8
WL	450	.4999933	.0701061	.265	.648
pythWL	450	.5005542	.0645627	.3024691	.6645963

Summary statistics are provided in Table 1 (above) and histograms for each variable are provided in the Data Appendix. The average turnover rate for a team is 31.65 percent with a standard deviation of 11 percent. On average, teams retained 64 percent of total games played from the previous season. It is interesting to note that, on average teams retained games played at catcher, shortstop, and starting pitcher at higher rate than for the entire roster at (68 percent, 71 percent, and 72 percent respectively versus 64 percent for the entire roster). These statistics appear to support the findings of Glenn, McGarrity, and Weller (2001). Therefore, they justify the inclusion of the positional-level secondary explanatory variables. On average, teams have 17 significant players. However, there is substantial variation from observation to observation, as there is a standard deviation of 2.96 and a range from 4 to 25 significant players on a roster. It is no surprise that this variation is also exhibited in the proportion of significant players on a roster, which ranges from .15 to .89.

However, the WAR associated with the average significant player is much more stable, averaging 1.875, and only ranging from .16 to 3.65, with a standard deviation of .548.

On average, teams have only 1.90 significant players who played in the opposite league in the previous season, a low number that appears to suggest teams actually prefer to avoid inter-league turnover. Finally, as expected, the average winning percentage is approximately equal to 50 percent. However, interestingly, the expected winning percentage calculated using the Pythagorean Expectation Formula projects a slightly higher average winning percentage with a smaller range and lower standard deviation than actual winning percentage.

VI. METHODOLOGY

A) Model of General Turnover Effects:

I begin my analysis by testing the hypothesis that turnover has a significant effect on team performance. I have previously explained the theory behind the prediction that there is an optimal turnover rate, meaning that the relationship between performance and turnover is non-linear. First, I test this hypothesis using a quadratic model. To quantify performance, I use *WL* (winning percentage) as opposed to *W* (*Wins*) as my dependent variable because teams do not always play the exact same number of games in a season. For example, some teams will only play 161 games in a season as opposed to the schedule 162 games due to irrelevant game cancellations. Therefore, winning percentage is a more consistent measure of team performance and serves as an extremely accurate proxy for firm productivity. I use *turnover* as my main explanatory variable. For year *t* and team *i*, this quadratic model is specified as:

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 turnover_{t,i}^2 + B_3 man_tenure_{t,i} + B_4 warpersig_{t,i} + B_5 sigrate_{t,i} + B_6 log(teampayroll)_{t,i} + B_7 WL_{t-1,i} + e_t$$

Next, I run an OLS regression that tests for the linear relationship between turnover and performance found by Seibert and Zubanov (2009). The model is specified without the quadratic $turnover_{t,i}^2$ term:

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 log(teampayroll)_{t,i} + B_6 WL_{t-1,i} + e_t$$

I also test whether winning percentage in the previous season influences the effect of turnover in the next season. To do so, I add an interaction term, $turnover * WL_{t-1}$. It is plausible that teams with worse winning percentages improve with turnover as it is easier to increase roster quality. Therefore, they would utilize job matching theory. It is also plausible that teams

with better winning percentages get worse with turnover because changes disrupt a team that is already performing well. In this case, FSHCM could be applied. It is important to note that variables used in interaction terms are centered throughout this paper, but only in models that include the interaction term. This transformation is done to limit the multicollinearity between the main effect and interaction within the model (Robinson and Schumacker, 2009). For robustness, I also run regressions that control for team fixed effects, due to the possibility that there is unobserved heterogeneity between teams.

In testing whether offseason turnover has an effect on team performance, it is necessary to control for manager tenure because, *ceteris paribus*, a manager with more experience with a team may improve that team's performance. I control for roster quality amongst significant players using *warpersig*. The inclusion of this variable in my model allows me to control for the quality of the significant players on a team's Opening Day roster. A team with better significant players will perform better than teams whose significant players are of lower quality. *Sigrate* controls for the number of significant players on the Opening Day Roster. A team with more significant players will perform worse than teams with less significant players because these teams will generally have more experienced players on the field. I control for WAR per significant player and the proportion of significant players instead of the sum of WAR represented in the variable *OpenWAR* because the latter variable is highly correlated with turnover. The product of *warpersig* and *sigrate* is proportional to *OpenWAR*. By dividing this variable into variables for WAR per player and the proportion of significant players, I eliminate the multicollinearity issue. Although, each of these variables is slightly correlated with turnover, they do not lead to multicollinearity within the model when separated.

In addition to controlling for player quality based on performance in the previous season, I control for player quality based on career performance through *team payroll*. A player with a higher salary has likely earned those wages through high expected productivity over multiple seasons. That player's performance over only the previous season may not be reflective of their potential performance in the next season. In other words, teams with higher payrolls may perform better over a season, even after controlling for roster quality based on player performance over the previous season. The log of payroll is taken because the relationship between payroll and team performance exhibits the law of diminishing returns.

Finally, it is possible that, *ceteris paribus*, team culture and mentality based on team performance in the previous season may affect team performance in the next season. Teams engrained with the confidence of past success — a winning culture — may perform better regardless of roster quality. Therefore, I control for WL_{t-1} , or team winning percentage over the previous season.

With these controls in place, a statistically significant, positive coefficient for turnover in the OLS model ($B_1 > 0$) would provide evidence for the hypothesis that job matching theory explains the relationship between turnover and performance in Major League Baseball. A statistically significant, negative coefficient for turnover ($B_1 < 0$) would provide evidence for the hypothesis that FSHCM explains the relationship between turnover and performance in Major League Baseball. A statistically significant, positive coefficient for turnover, coupled with a statistically significant, negative coefficient for turnover² would provide evidence to support the hypothesis that there is an optimal point of turnover for Major League Baseball teams. In this

case, job matching theory would dictate at low turnover rates, and FSHCM would dictate at high turnover rates.

B) Secondary Models of Turnover:

Next, I conduct a series of OLS regressions based on my hypotheses regarding where I would expect to see job matching theory and FSHCM manifest itself in baseball. Previously, I identified three areas where these theories could function. FSHCM could operate at the positional level, and job matching theory could apply to turnover across leagues. Either theory could apply to the effect of ballpark “quirkiness” on turnover.

In order to test for the effect of positional turnover on team performance, I replace *turnover* with three separate explanatory variables: *catchercont*, *sscont*, *startpitchcont*. These variables represent the inverses of catcher, shortstop, and starting pitcher turnover respectively. The positional turnover regression is as follows:

$$WL_{t,i} = B_0 + B_1 catchercont_{t,i} + B_2 sscont_{t,i} + B_3 startpitchcont_{t,i} + B_4 man_tenure_{t,i} + B_5 warpersig_{t,i} + B_6 sigrate_{t,i} + B_7 \log(teampayroll)_{t,i} + B_8 WL_{t-1,i} + e_t$$

Although I run a two-sided test, my prediction framing this regression is that $B_1, B_2, B_3 > 0$. If there is evidence to support this hypothesis, that means that there is evidence that continuity, or a lack of turnover, at positions that require the most communication perform better. As previously explained, this conclusion would provide evidence that FSHCM is operating at the positional level in Major League Baseball.

In order to test for the effects of inter-league turnover, I replace *turnover* with *leagueturnover*, which represents the number of players on each team who played in the opposite league the previous season. The league turnover regression is as follows:

$$WL_{t,i} = B_0 + B_1 leagueturnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 \log(teampayroll)_{t,i} + B_6 WL_{t-1,i} + e_t$$

Although I run a two-sided test, the prediction providing the basis for this regression is that $B_1 > 0$. If there is evidence to support this hypothesis, that means that there is evidence that teams that get more players from the other league are using the principles of job matching theory to improve their performance.

Finally, in order to test for whether a team's ballpark influences whether job matching theory or FSHCM can explain the effect of turnover on performance, I include *parkfactor* in the original model, as well as an interaction term representing the product of *parkfactor* and *turnover*. The park factor regression is as follows:

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 parkfactor + B_3 (turnover * parkfactor) + B_4 man_tenure_{t,i} + B_5 warpersig_{t,i} + B_6 sigrate_{t,i} + B_7 \log(teampayroll)_{t,i} + B_8 WL_{t-1,i} + e_t$$

It is necessary to include an interaction term between park factor and turnover because the theory behind the model is that teams with higher park factors will be affected differently by turnover. Therefore, if there is evidence to support the hypothesis that $B_3 > 0$, then there is evidence to support the notion that teams with quirkier ballparks can optimize their performance by applying the principles of job matching theory. If $B_3 < 0$, then teams utilize FSHCM. For robustness, these generalized models are combined together to include all secondary explanatory variables as well as incorporated into the original model if necessary. All tests are run with and without team fixed effects.

C): Alternative Measures of Key Variables

In order to both illustrate the robustness of my results, and to test for alternative hypotheses, I conduct my general turnover regressions using a number of alternative measures. First, I replace my main explanatory variable, the turnover of significant players, with roster continuity. Roster continuity measures turnover at the game level, and includes every player, not just those that I have deemed significant. Using this model allows me to verify that my results have not been influenced by the measure of turnover previously used. The adjusted model is as follows:

$$WL_{t,i} = B_0 + B_1 rostercont_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 \log(teampayroll)_{t,i} + B_6 WL_{t-1,i} + e_t$$

Secondly, I use an alternative measure of performance in the previous season — Pythagorean Winning Percentage — calculated using the formula developed by Bill James. Using this measure controls for the possibility that a team's winning percentage in the previous season did not reflect their actual performance. The adjusted model is as follows:

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 \log(teampayroll)_{t,i} + B_6 pythWL_{t-1,i} + e_t$$

Thirdly, I use the Las Vegas Over/Under win projections, which are based on baseball projection systems and the quality of the players on the roster, to control for overall team quality. Because the Vegas projections are highly correlated with *OpenWAR*, *WL_{t-1}* and, *Team Payroll*, it is not necessary to include any of the corresponding explanatory variables in a regression that also includes *vegasW*. Using this measure allows me to verify the robustness of my primary measures of team quality. The adjusted model is as follows:

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 vegasW_t$$

Finally, I replace my dependent variable, winning percentage, with a measure of short term performance. I decide to look at performance up until the All Star Break because it usually reflects a team's performance over about half the season. Furthermore, it comes about two weeks before the July 31st trade deadline, around which time teams will make their most significant in-season roster moves. By looking at performance only up until that point, I make sure that I am not conflating my analysis of the effect of offseason turnover with the effect of the most significant in-season turnover. The adjusted model is as follows:

$$AllStarWL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 \log(team payroll)_{t,i} + B_6 WL_{t-1,i} + e_t$$

VII. RESULTS

A) Models of General Turnover:

Table 2: Quadratic Estimate of the Effect of General Turnover on Win %

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Turnover	.0233 (.1056)	.0438 (.0949)	.0539 (.0938)	-.0110 (.0933)	-.0079 (.0916)	-.0224 (.0964)
Turnover²	-.1529 (.1473)	-.1381 (.1324)	-.09411 (.1315)	.0167 (.1317)	.0010 (.1293)	.0077 (.1360)
Manager Tenure	.0042*** (.0009)	.0022*** (.0008)	.0019** (.0008)	.0014* (.0008)	.0008 (.0008)	-.0008 (.0011)
WAR per Significant Player		.0550*** (.0055)	.0529*** (.0055)	.0428*** (.0059)	.0255*** (.0071)	.0311*** (.0079)
Prop. of Significant Players			.0977*** (.0293)	.0765*** (.0292)	.0363 (.0303)	.0584* (.0320)
Log(Team Payroll)				.0305*** (.0074)	.0259*** (.0073)	.0258*** (.0010)
Win %_{t-1}					.2435*** (.0596)	.1366** (.0655)
Team Fixed Effects	No	No	No	No	No	Yes
Adjusted R² (Prob>F)	.0662 (.0000)	.2446 (.0000)	.2625 (.0000)	.2903 (.0000)	.3163 (.0000)	.3133 (.0000)

***=Statistically significant at 99% level, **=Statistically Significant at 95% Level, *=Statistically Significant at 90% Level

Unlike the results Harris, Tang, and Tseng (2006) found when looking at Australian Firms, there is absolutely no evidence to suggest that there is an optimal point of turnover that Major League Baseball teams should target, as can be seen in Table 2. Furthermore, as can be seen in Table 3, there is weak to no evidence of a linear relationship between turnover and performance, as Seibert and Zubanov (2009) found when looking at full-time employees for UK retail companies. Each table presents the coefficients of the independent variables used in the regression to estimate a team's win percentage over a full season, with standard errors in

parentheses. Both the basic quadratic and OLS models are presented by adding controls one by one. An interaction term is added to the OLS estimate in Table 3, and team fixed effects results are included in both tables. As can be seen in Table 2, turnover is never statistically significant in the quadratic model, even with only manager tenure used as an additional control.

Table 3: OLS estimate of the Effect of General Turnover on Win %

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Turnover	.0824*** (.0280)	-.0516** (.0253)	-.0102 (.0278)	.0003 (.0273)	-.0073 (.0269)	-.017 (.0285)	-.0035 (.0270)	-.0128 (.0287)
Manager Tenure	.0043*** (.0009)	.0022*** (.0008)	.0020** (.0008)	.0014* (.0008)	.0008 (.0008)	-.0008 (.0011)	.0008 (.0008)	-.0008 (.0011)
WAR per Significant Player		.0551*** (.0055)	.0529*** (.0055)	.0429*** (.0059)	.0255*** (.0071)	.0311*** (.0079)	.0253*** (.0071)	.0306*** (.0079)
Prop. of Significant Players			.0998*** (.0292)	.0763*** (.0291)	.0362 (.0302)	.0583* (.0319)	.0350 (.0302)	.0571* (.0319)
Log(Team Payroll)				.0304*** (.0072)	.0258*** (.0071)	.0257*** (.0098)	.0252*** (.0072)	.0242** (.0098)
Win %_{t-1}					.2436*** (.0595)	.1367** (.0653)	.2418*** (.0595)	.1371** (.0652)
Turnover*Win%_{t-1}							.4452 (.325)	.4410 (.3450)
Team Fixed Effects	No	No	No	No	No	Yes	No	Yes
Adjusted R² (Prob>F)	.0662 (.0000)	.2444 (.0000)	.2634 (.0000)	.2920 (.0000)	.3179 (.0000)	.3133 (.0000)	.3194 (.0000)	.3163 (.0000)

***=Statistically significant at 99% level, **=Statistically Significant at 95% Level, *=Statistically Significant at 90% Level

However, Table 3 illustrates that, in OLS Models 1 and 2, turnover is significant at the 99 percent and 95 percent levels respectively. However, particularly in OLS Model 1, the low adjusted r^2 indicates that the model is under-specified. On the other hand, in OLS Model 2, turnover is still significant at the 95 percent level, even when controlling for the quality of

significant players. Furthermore, the sign has flipped from Model 1. According to this model, a 20 percent increase in turnover will lead to 1 percent decrease in winning percentage, or 1.62 wins in a season. Manager tenure and WAR per significant player on the Opening Day roster are also significant, but at the 1 percent level. A one-unit increase in WAR per significant player will increase a team's winning percentage by .05, or about 8 wins. A manager tenured for 10 years will help the team earn 3.24 wins compared to a newly hired manager. However, even though the r^2 of OLS Model 2 and the fully specified OLS Model 5 are somewhat comparable, the 25 percent increase in explanatory power of Model 5 indicates that Model 2 is under-specified and therefore any evidence of a negative effect of turnover on team performance, which would be explained by FSHCM, is weak at best.

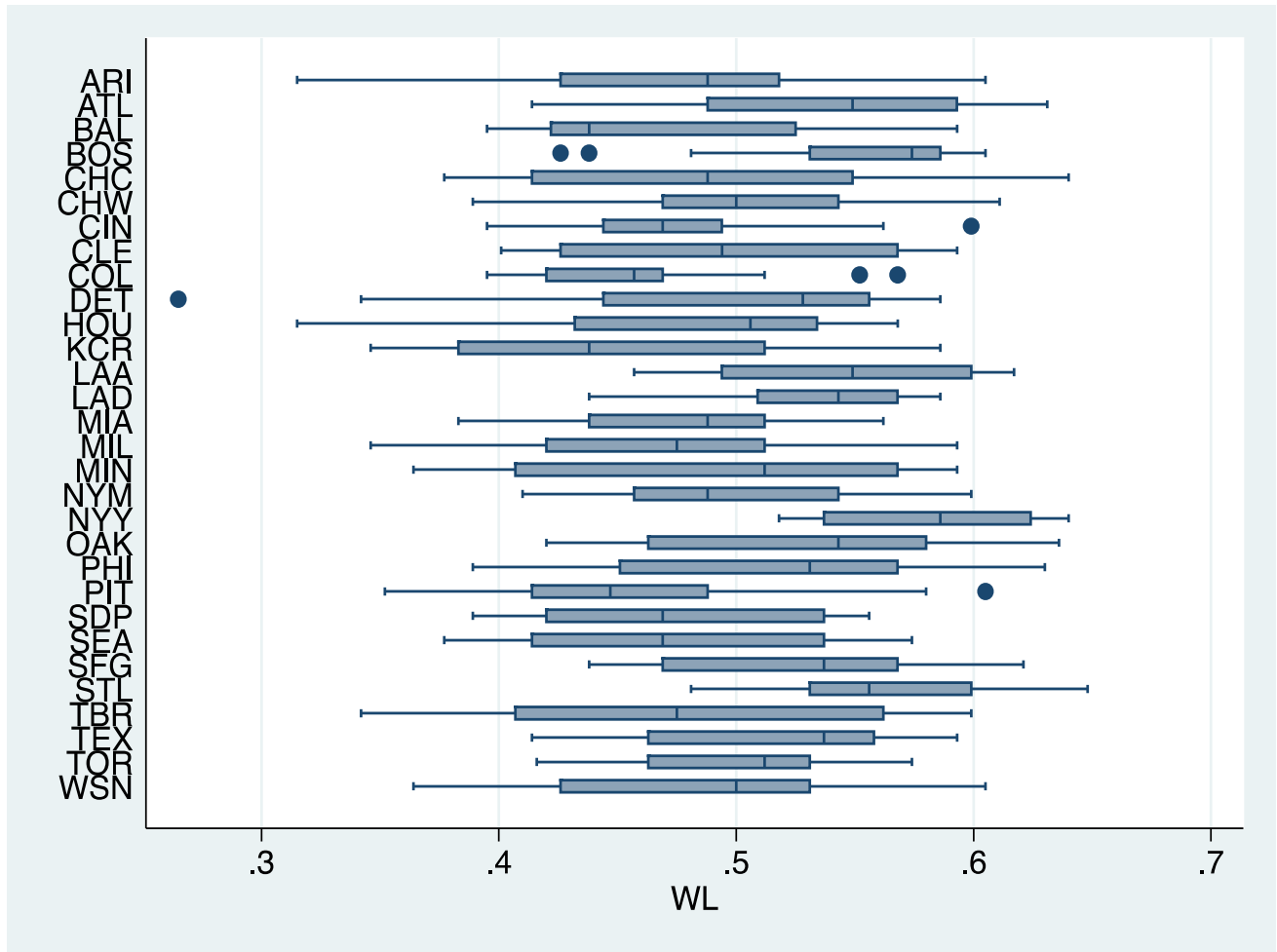
In both the quadratic and OLS fully specified models excluding team fixed effects, (Model 5 in Tables 2 and 3 respectively), the coefficient associated with turnover is not statistically significant. Furthermore, the interaction term between turnover and winning percentage in the previous season, introduced in Models 7 and 8 of Table 3, is not statistically significant. However, the coefficients of the remaining explanatory variables are essentially equivalent across the OLS and quadratic models. In addition to turnover, manager tenure and the proportion of significant players making up a team's Opening Day roster are not statistically significant. However, a unit increase in WAR per significant player is predicted to lead to a .0255 increase in winning percentage, or about 4 wins over the course of the regular season. One way of interpreting this result is to base it off the mean number of significant players on a team. On average, teams have about 17 significant players on their Opening Day Rosters. For WAR per significant player to increase by one unit, a team must add 17 WAR. In this average case, adding

17 WAR will lead to just over 4 wins over the course of the season, or only 25 percent of expected return on average.

The log of team payroll is also statistically significant at the 99 percent level. However, with roster quality held constant, a 20 percent increase in payroll leads to only a tiny increase in winning percentage equivalent to less than half of a game. Finally, winning percentage from the previous season is also significant at the 99 percent level. A 10 percent increase in winning percentage in the previous season will increase a team's winning percentage in the next season by 2.5 percent in the next season, or by about 4 wins. Another way of interpreting this coefficient is that any positive change in a team's winning percentage in the previous season leads to a positive increase of in winning percentage in the next season of about 25 percent of that change.

When including team fixed effects in Model 6 of each table, it is important to note that the significance of winning percentage from the previous season decreases from the 99 percent to the 95 percent level. In this model, any positive change in a team's winning percentage from the previous season leads to a positive increase in winning percentage in the next season of about 13 percent of that change. This effect is about half the size specified in the model without team fixed effects. It is plausible that this difference is the result of team controls capturing some of the impact of past performance. Evidence supporting this explanation can be seen in Graph 1. Certain teams, such as the New York Yankees and the Los Angeles Dodgers, have systematically performed better from 2002-2016, and certain teams, such as the Pittsburgh Pirates and the Cincinnati Reds, have performed worse.

Graph 1: Win % Distribution by Team



It should also be noted that the proportion of significant players becomes significant at the 90 percent level. A 20 percent increase in the proportion of significant players, or 5 more significant players on a 25-man roster, increases winning percentage by .01, or 1.62 wins per year. Additionally, the positive effect of WAR per significant player increases by about 20 percent. With individual franchise quality controlled for, the statistical significance of significant player quality and quantity increases.

Numerous robustness checks were conducted to confirm the validity of these results. Variance inflation factor (VIF) tests were performed to test for multicollinearity issues. A VIF test determines the extent to which multicollinearity decreases the accuracy of the estimates in a model (Robinson and Schumacker, 2009) by regressing each independent variable against the remaining explanatory variables in the model and recording the coefficient of determination.

For each variable, $VIF = \frac{1}{1-R^2}$

A variable with a VIF greater than 5 does not have an accurate coefficient estimate due to multicollinearity. No evidence of multicollinearity was found through these tests and results for the paper's main models are displayed in the Robustness Appendix.

A Breusch-Pagan test for heteroskedasticity (Breusch and Pagan, 1979) was conducted based on values of the fitted regression. This robustness test is a simple chi-squared test that provides a probability that the residuals are independent of the explanatory variables in the regression. A p-value less than .05 provides evidence of heteroskedasticity. No evidence of heteroskedasticity was found using these robustness tests and results for the paper's main models are also displayed in the Robustness Appendix. With these robustness tests complete, it is safe to conclude that there is weak to no evidence of FSHCM explaining a team's performance over a full season through turnover, and no evidence of the turnover effects of job matching theory on performance over a full season.

B) Secondary Models of Turnover:

Table 4: OLS Estimates of the Effects of Positional Turnover on Win %

	Model 1	Model 2	Model 3	Model 4	Model 5
Catcher Continuity	.0077 (.0103)	.0059 (.0101)			.0074 (.0107)
Shortstop Continuity	-.0058 (.0087)		-.0047 (.0086)		-.0046 (.0093)
Starting Pitcher Continuity	-.0146 (.0152)			-.0134 (.0151)	-.0117 (.0160)
Manager Tenure	.0008 (.0008)	.0008 (.0008)	.0008 (.0008)	.0009 (.0008)	-.0007 (.0011)
WAR per Significant Player	.0257*** (.0072)	.0254*** (.0071)	.0256*** (.0071)	.0261*** (.0059)	.0313*** (.0080)
Prop. of Significant Players	.0454 (.0288)	.0372 (.0279)	.0415 (.0278)	.0455 (.0284)	.0698** (.0304)
Log(Team Payroll)	.0260*** (.0071)	.0258*** (.0071)	.0263*** (.0071)	.0259*** (.0071)	.0267*** (.0098)
Win %_{t-1}	.2439*** (.0595)	.2418*** (.0594)	.2436*** (.0594)	.2433*** (.0594)	.1378** (.0655)
Team Fixed Effects	No	No	No	No	Yes
Adjusted R² (Prob>F)	.3173 (.0000)	.3184 (.0000)	.3183 (.0000)	.3289 (.0000)	.3171 (.0000)

***=Statistically significant at 99% level, **=Statistically Significant at 95% Level, *=Statistically Significant at 90% Level

With no strong evidence to support my main hypothesis that a combination of job matching theory and FSHCM explains the effect of general turnover on team performance in Major League Baseball, it is important to test any extensions of this hypothesis to ensure that neither theory is operating on a micro-scale. As can be seen in Table 4, there is no evidence that turnover at catcher, shortstop, and pitcher — positions that require substantial communication

— negatively affect team performance. Therefore, there is no evidence to support the hypothesis that FSHCM manifests itself through positional turnover in Major League Baseball.

Table 5: OLS Estimates of the Effects of Park Factors and League Turnover on Win %

	Model 1	Model 2	Model 3	Model 4
League Turnover	-.0004 (.0021)		-.0005 (.0020)	-.0008 (.0021)
Turnover*Park Factor	-.0004 (.0080)	-.0005 (.0080)		
Manager Tenure	.0008 (.0008)	.0008 (.0008)	.0008 (.0008)	-.0008 (.0011)
WAR per Significant Player	.0257*** (.0072)	.0256*** (.0072)	.0257*** (.0071)	.0312 (.0079)
Prop. of Significant Players	.0377 (.0328)	.0355 (.0304)	.0415 (.0287)	.0694 (.0308)
Log(Team Payroll)	.0259*** (.0072)	.0259*** (.0072)	.0260*** (.0071)	.0263 (.0097)
Win %_{t-1}	.2405*** (.0605)	.2421*** (.0598)	.2407*** (.0598)	.1327 (.0656)
Team Fixed Effects	No	No	No	Yes
Adjusted R² (Prob>F)	.3173 (.0000)	.3134 (.0000)	.3179 (.0000)	.3138 (.0000)

***=Statistically significant at 99% level, **=Statistically Significant at 95% Level, *=Statistically Significant at 90% Level

Furthermore, Table 5 illustrates that there is no evidence to support the hypothesis that inter-league turnover affects team performance. There is also no evidence to support the hypothesis that a team's ballpark changes the relationship between turnover and performance. Therefore, there is no evidence to suggest that either FSHCM or job matching theory operates through league turnover or the quirkiness of a team's ballpark in Major League Baseball.

VIF tests for multicollinearity and Breusch-Pagan tests for heteroskedasticity were conducted for every model listed, and there was never any evidence of either problem. Also, it

is important to note the consistency exhibited in the estimates of control coefficients in these models. The interpretations from the previous section for the coefficients for manager tenure, WAR per significant player, the proportion of significant players, team payroll, and winning percentage from the previous season do not change, regardless of the turnover specification used. These results indicate the robustness of these coefficients.

C) Alternative Measures of Key Variables

Table 6: Measuring Turnover Using Roster Continuity

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Roster Continuity	.1392 (.2343)	.1893 (.2436)	-.0249 (.0306)	-.0018 (.0313)	-.0238 (.0307)	-.0013 (.0342)
Roster Continuity²	-.1308 (.1852)	-.1534 (.0011)				
Manager Tenure	.0008 (.0008)	-.0009 (.0011)	.0008 (.0008)	-.0008 (.0011)	.0008 (.0008)	-.0008 (.0011)
WAR per Significant Player	.0263*** (.0072)	.0317*** (.0080)	.0261*** (.0072)	.0313*** (.0079)	.0261*** (.0072)	.0313*** (.0080)
Prop. of Significant Players	.0452 (.0292)	.0645** (.0307)	.0468 (.0291)	.0663** (.0305)	.0466 (.0293)	.0657* (.0308)
Log(Team Payroll)	.0256*** (.0072)	.0257*** (.0099)	.0264*** (.0071)	.0266*** (.0098)	.0264*** (.0072)	.0263*** (.0010)
Win %_{t-1}	.2523*** (.0602)	.1410** (.0664)	.2473*** (.0597)	.1356** (.0661)	.2470*** (.0599)	.1348** (.0663)
Continuity*Win%_{t-1}					-.0329 (.3966)	-.0808 (.4172)
Team Fixed Effects	No	Yes	No	Yes	No	Yes
Adjusted R² (Prob>F)	.3180 (.0000)	.3139 (.0000)	.3188 (.0000)	.3138 (.0000)	.3172 (.0000)	.3133 (.0000)

As can be seen in Table 6, using *roster continuity* as an alternative measure for turnover — a measure that is at the game level and includes all players — does not alter the conclusions that can be drawn from the model. The coefficients and significance levels of the remaining variables remain stable when comparing these results to those in Tables 2 and 3. Furthermore, *roster continuity* is never statistically significant when the model is fully specified, just as was found when looking at *turnover*.

Table 7: Replacing Win % $t-1$ with Pythagorean Expectation Win % $t-1$

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Turnover	-.0065 (.0919)	-.01934 (.0967)	-.0027 (.0269)	-.0139 (.0286)	.0005 (.0271)	-.0097 (.0288)
Turnover²	.0055 (.1297)	.0081 (.1363)				
Manager Tenure	.0010 (.0008)	-.0008 (.0011)	.0010 (.0008)	-.0008 (.0011)	.0009 (.0008)	-.0008 (.0011)
WAR per Significant Player	.0261*** (.0072)	.0324*** (.0081)	.0261*** (.0073)	.0324*** (.0081)	.0262*** (.0073)	.0321*** (.0081)
Prop. of Significant Players	.0410 (.0292)	.0628** (.0319)	.0409 (.0302)	.0627** (.0318)	.0407 (.0302)	.0626** (.0318)
Log(Team Payroll)	.0276*** (.0072)	.0276*** (.0099)	.0275*** (.0071)	.0275*** (.0097)	.0270*** (.0071)	.0260*** (.0098)
Pythagorean Win%$_{t-1}$.2434*** (.0602)	.1244* (.0715)	.2434*** (.0648)	.1246* (.0714)	.2405*** (.0649)	.1238* (.0713)
Turnover*Pythagorean Win%$_{t-1}$.3825 (.3505)	-.4166 (.3698)
Team Fixed Effects	No	Yes	No	Yes	No	Yes
Adjusted R² (Prob>F)	.3121 (.0000)	.3075 (.0000)	.3137 (.0000)	.3075 (.0000)	.3140 (.0000)	.3092 (.0000)

As can be seen in Table 7, using *Pythagorean Win %* as an alternative measure for *Win % $_{t-1}$* — a measure that bases past performance on runs scored and runs allowed — does not alter the conclusions that can be drawn from the model. The coefficients and significance levels of the remaining variables also remain stable when comparing these results to those in Tables 2 and 3. It is interesting to note that the alternative measure of past performance consistently had slightly lower, yet still statistically significant, coefficients. Even though these differences are fractional, the results seem to corroborate my suspicions that the Pythagorean Expectations Formula is not as effective when used across multiple seasons.

Table 8: Using Vegas Over/Under Win Predictions to control for Team Quality

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Turnover	-.0637 (.0901)	-.0820 (.0952)	-.0060 (.0246)	-.0110 (.0263)	-.0065 (.0251)	-.0101 (.0270)
Turnover²	.0839 (.1262)	.1035 (.1334)				
Manager Tenure	.0010 (.0008)	-.0007 (.0011)	.0010 (.0008)	-.0008 (.0011)	.0010 (.0008)	-.0008 (.0011)
Vegas Over/Under	.0049*** (.0004)	.0044*** (.0005)	.0049*** (.0004)	.0044*** (.0005)	.0049*** (.0004)	.0043*** (.0029)
Turnover*Vegas Over/Under					-.0002 (.0027)	.0004 (.0029)
Team Fixed Effects	No	Yes	No	Yes	No	Yes
Adjusted R² (Prob>F)	.3302 (.0000)	.3276 (.0000)	.3312 (.0000)	.3269 (.0000)	.3295 (.0000)	.3267 (.0000)

As can be seen in Table 8, using *Vegas Over/Under* as an alternative measure of controlling for team quality does affect the coefficients of turnover and manager tenure. In fact, the new variable appears to be the source of most of the explanatory power in the model. Overall, using the alternative control measures of *roster continuity*, *Pythagorean Win %*, and *Vegas Over/Under* appear to demonstrate the robustness of the independent variables used to estimate the impact of roster turnover on team performance over a full season.

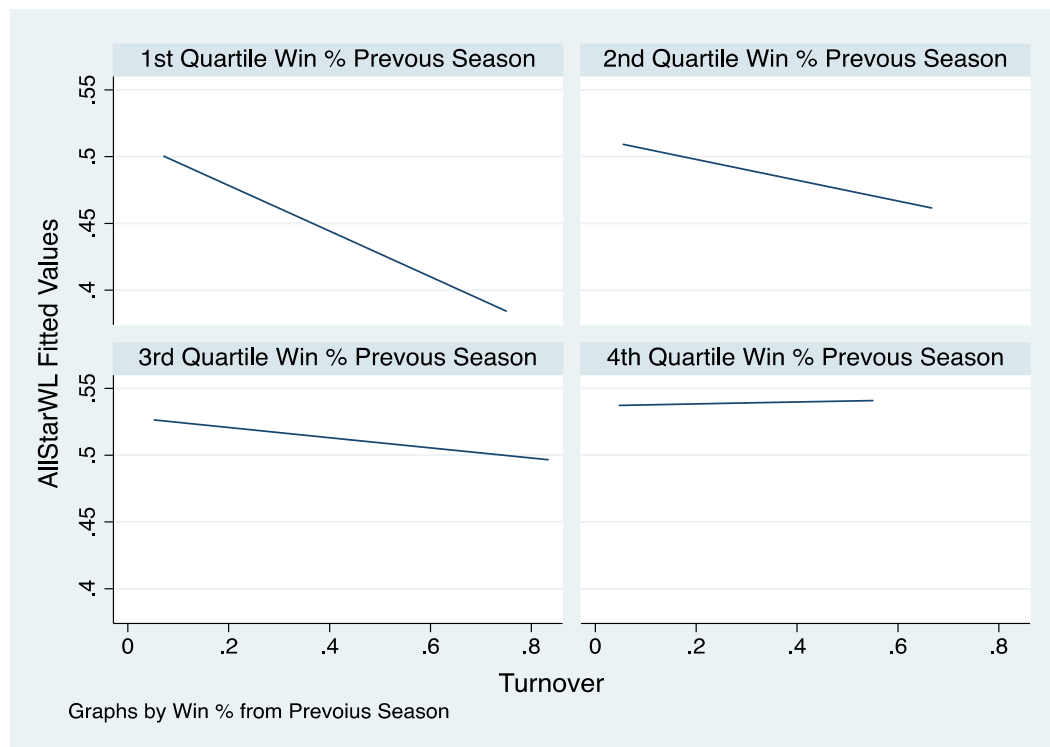
Table 9: Using All Star Break Win % as Dependent Variable

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Turnover	-.0333 (.0978)	-.0283 (.1027)	-.0351 (.0003)	-.0395 (.0304)	-.0294 (.0288)	-.0324 (.0305)
Turnover²	.0028 (.1381)	-.0166 (.1448)				
Manager Tenure	.0002 (.0009)	-.0013 (.0012)	.0003 (.0009)	-.0013 (.0012)	.0002 (.0009)	-.0014 (.0012)
WAR per Significant Player	.0212*** (.0076)	.0247*** (.0085)	.0212*** (.0076)	.0247*** (.0084)	.0208*** (.0076)	.0238*** (.0084)
Prop. of Significant Players	.0261 (.0324)	.0452 (.0341)	.0262 (.0323)	.0455 (.0339)	.0243 (.0322)	.0436 (.0338)
Log(Team Payroll)	.0284*** (.0078)	.0248** (.0106)	.0284*** (.0076)	.0251** (.0104)	.0275*** (.0076)	.0227** (.0104)
Win%_{t-1}	.2568*** (.0637)	.1746** (.0697)	.2568*** (.0636)	.1743** (.0696)	.2540*** (.0634)	.1748** (.0693)
Turnover* Win%_{t-1}					.6911** (.3461)	.7040* (.3664)
Team Fixed Effects	No	Yes	No	Yes	No	Yes
Adjusted R² (Prob>F)	.2814 (.0000)	.2823 (.0000)	.2832 (.0000)	.2823 (.0000)	.2883 (.0000)	.2885 (.0000)

As can be seen in Table 9, when leaving the explanatory variables unchanged from the original model, and only changing the dependent variable to measure performance over the short term through the All Star Break, turnover interacted with past winning percentage becomes statistically significant. When the model does not include team fixed effects, the coefficient is statistically significant at the 95 percent level. With fixed effects, it is significant at the 90 percent level. However, the overall explanatory power of the model decreases by about 15 percent when attempting to predict short term performance. This result is not entirely surprising given that teams are more likely to over or under-perform relative to their quality

over a shorter time-frame. Based on the results, it appears that the direction of the relationship between turnover and short-run performance changes depending on the team's past winning percentage.

Graph 2: Fitted Values of All Star Win % vs. Turnover by Quartiles of Win %_{t-1}



This relationship is most clearly illustrated in Graph 2 above. As can be seen, the worst 25 percent of teams exhibit a strong negative relationship between turnover and performance up until the All Star break in the next season. The middle 50 percent of teams exhibit a weaker negative relationship. The top 25 percent of teams exhibit a weak positive relationship between turnover and performance in the next season. A 20 percent increase in turnover will lead to about 1 less win over the first half of the season for a team that won 70 games, or had a winning percentage of .432. One win can be substantial over the course of a season. A team with an average winning percentage of .500 in the previous season that increases turnover by

20 percent will win approximately .493 games less in the first half of the next season. However, a 20 percent increase in turnover for a team that won 90 games in the previous season (90 win teams will usually qualify for the playoffs) wins about .15 more games over the first half of the season. Unlike the negative effect of turnover for 70 win teams, this positive effect is not quantitatively significant.

Essentially, the negative effect of turnover on performance over the first half of the season diminishes the better a team was in the previous season. Poor teams experience the negative effect of turnover predicted by the FSHCM, an effect that decreases the better the team gets. The best teams actually experience a very small positive effect of turnover predicted by job matching theory.

VIII. DISCUSSION

According to my analysis, there is not sufficient evidence to confidently conclude that FSHCM and job matching theory apply to Major League Baseball as it relates to any potential relationship between turnover and team performance over a full season. However, there is evidence that turnover has an effect on performance over a shorter period of time. Over a half-season, better teams are not substantially affected by turnover, but worse teams are negatively affected. One possible explanation for this result may be that better teams are able to attract higher quality free agents who are attracted to the franchise by the prospect of winning. Therefore, those teams may be able to utilize job matching theory and find optimal replacements for any departing players. By doing so, they may be able to offset the costs of turnover associated with FSHCM. On the other hand, worse teams may struggle to replace players who depart because their franchises are not as attractive to players looking to move. Furthermore, it might be easier for better teams to make optimal trades because they have more high quality players to offer. Overall, this result appears to suggest that winning teams are more attractive transaction partners. Therefore, it appears that general managers of poor-performing teams should work hard to retain their best players because they will find it harder to replace them should they leave.

However, these conclusions must be placed in the context of the fact that the statistical significance of the interaction between turnover and past winning percentage disappears when looking at performance over a full season. It is plausible that there is a different relationship between mid-season turnover and performance that counteracts these short-term effects. Worse performing teams, given more time to replace their players, may be able to more

effectively do so, and start to improve over the second half of the season. Also, those teams that were unable to find already optimal matches may start to benefit from the accumulation of human capital that new players build up over the course of the season. Perhaps better teams who have already found optimal matches do not benefit from this improvement over time. Regardless, there is sufficient evidence to conclude that both FSHCM and job matching theory can be applied through the effects of turnover in Major League Baseball, at least in the short term. Better teams appear to be in a more favorable position to offset the costs of turnover associated with FSHCM with the positive effects associated with job matching theory. On the other hand, the worst teams must be wary of the negative effects of turnover outlined by FSHCM as they cannot as easily offset these costs.

However, there is no evidence to suggest that positional turnover, league turnover, or park factors have an effect on turnover. Nevertheless, it is possible that, in circumstances where no effect of turnover can be observed, that job matching theory and FSHCM are operating contemporaneously, and cancelling each other out. I have already explained how either theory could be used to explain the influence of ballpark effects on turnover. Teams with quirkier ballparks may benefit from turnover through job matching theory because they are able to find players whose skillsets are optimal to the home stadium. However, it is equally possible that FSHCM explains this relationship. Players may need time to adjust to quirkier ballparks. Therefore, it is an equally valid prediction to say that higher rates of turnover could negatively affect performance among those teams. When considering general turnover, it is possible that the benefits of turnover associated with job matching and the costs of turnover associated with FSHCM are equal and simply cancel each other out over a full season.

Beyond the fact that the turnover of significant players may have a different effect on short-term performance depending on how good the team was in the previous season, my model provides more interesting information through its controls. General managers are most concerned with their team's performance over the upcoming full season. Given that turnover is not a significant predictor of how well their teams will do over that time period, and given the that an increase in the quality of significant players on the roster increases performance, it appears that team executives should be more concerned with improving the quality of their rosters than with maintaining chemistry or communication, particularly if their team has already performed well in the past.

Furthermore, the statistical and quantitative significance of win percentage from the previous season appears to indicate that a winning culture breeds more winning. If a team can find a way to win games then, regardless of roster quality, they are more likely to win games in the next season. Part of that relationship appears to be that better teams have a greater capacity to attract better players. Therefore, it may be advisable for general managers to act in the best interest of their team in the short-term to breed a winning culture and make their franchise more appealing to transaction partners, rather than fostering a losing culture by focusing on the development of inexperienced or lesser skilled players.

Even though my analysis was able to yield some interesting information for general managers, there were some limitations to my ability to analyze the impact of turnover on team performance as accurately as possible. Without the resources of a large staff, developing and running a state of the art baseball projection system that is reproducible and transparent is not possible within the timeframe of a year-long thesis. As Nate Silver, the inventor of PECOTA, says

of his projection system, “It’s my baby, but it takes a village to run.”¹ Incorporating the data from a baseball projection system into preseason roster quality controls could be beneficial to future research. Furthermore, because the relationship between turnover and performance over the short term differs from the relationship over a full season, it may have been useful to control for mid-season factors such as injuries and in-season acquisitions. Future research could incorporate these factors into their models as well.

¹ Hageman, William. 2006. “Baseball by the Numbers.” Retrieved April 1, 2017, from http://articles.chicagotribune.com/2006-01-04/features/0601040237_1_major-league-korean-league-past-players.

X. CONCLUSION

The goal of this research was to determine if turnover affected team performance in Major League Baseball, and to assess whether job matching theory or the Firm Specific Capital Model could explain this relationship. My results provide weak to no evidence of any effect of turnover on performance over a full season. Any weak evidence points in the direction of FSHCM: that there is a negative relationship between turnover and winning percentage. However, over only half a season, better teams benefit from turnover through job matching theory, and worse teams are harmed through FSHCM. These results indicate that better teams find it easier to attract better players, and it takes time for worse teams to find effective replacements for any significant players they lose.

Beyond this general analysis, I expanded on past analysis of these two models by attempting to pinpoint specific areas where either FSHCM or job matching theory applied in the industry in question. I could not find any evidence that FSHCM could be applied to positional turnover, or that job matching theory could be applied to inter-league turnover or to ballpark characteristics. However, the quality of the best players on the roster, and winning percentage from the previous season were shown to have a positive effect on team performance. This conclusion may seem rather intuitive. However, it provides a general roadmap to success for general managers, based on the notion that the acquisition of talent is the best route to improved performance. Better talent leads to more winning, which leads to even more winning. Given that offseason turnover does not significantly impact team performance over a full season, factors such as a player's effect on chemistry, communication, or his suitability to the

team, should remain secondary to recent performance when assessing whether to acquire a new player.

Data Appendix

Unit of Observation for Analysis Dataset: Team/ Year

Range of Years: 2002-2016

Number of Teams: 30

Coding of Teams:

Value	Meaning	Value	Meaning
1	ARI	16	MIL
2	ATL	17	MIN
3	BAL	18	NYM
4	BOS	19	NYN
5	CHC	20	OAK
6	CHW	21	PHI
7	CIN	22	PIT
8	CLE	23	SDP
9	COL	24	SEA
10	DET	25	SFG
11	HOU	26	STL
12	KCR	27	TBR
13	LAA	28	TEX
14	LAD	29	TOR
15	MIA	30	WSN

Variable Name: turnover

Original Data Source: Player Value Data by Year Hitters, Player Value Data By Year Pitchers

Missing Observations: 30/450

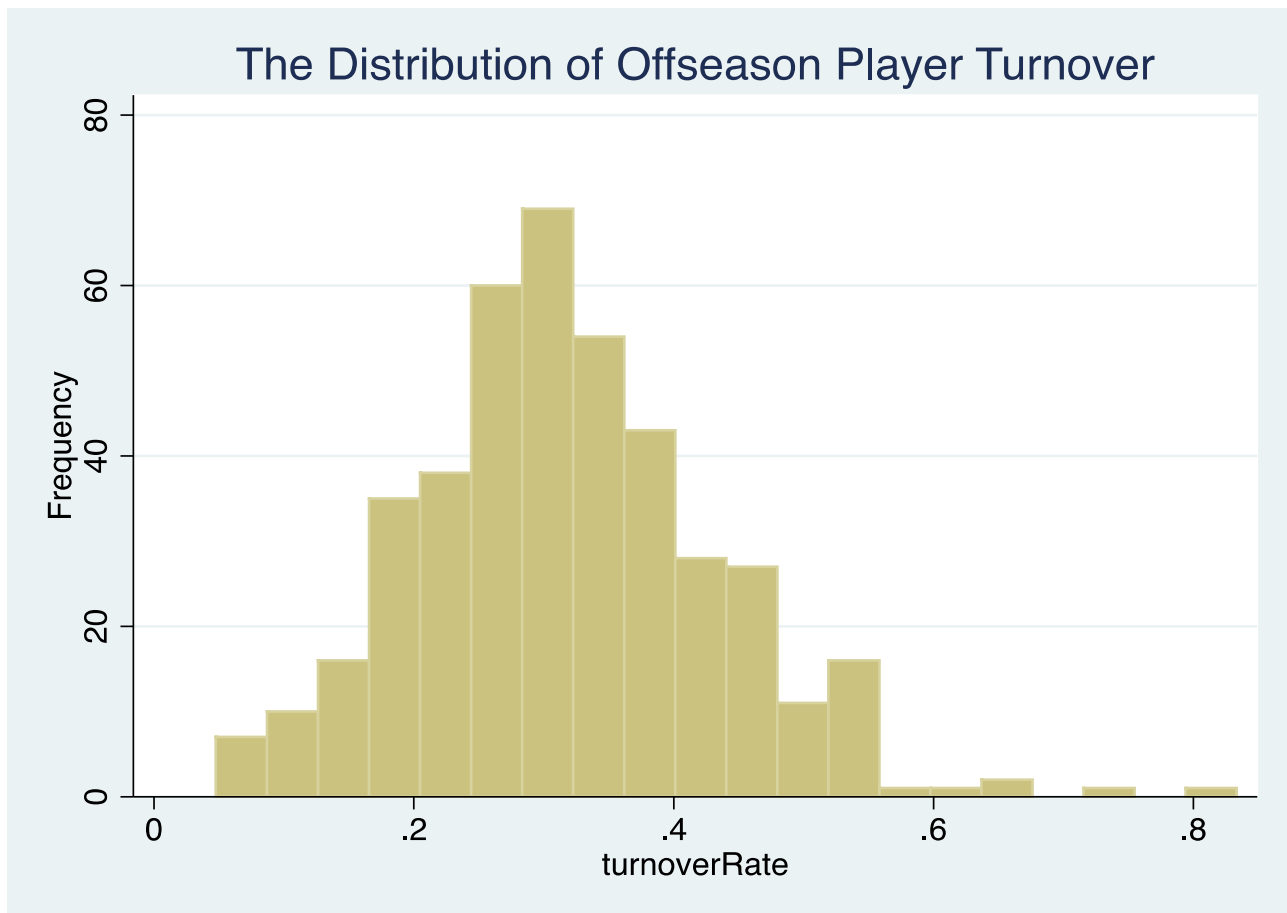
Definition: The proportion of significant players from the previous season who are not on the opening day roster.

Units: N/A

Mean=.316503 **Standard Deviation=**.115693

Min=.048 **25th percentile=**.238 **Median=**.304 **75th percentile=**.381 **Max=**.833

Histogram:



Variable Name: rostercont

Original Data Source: Rosters

Missing Observations: 30/450

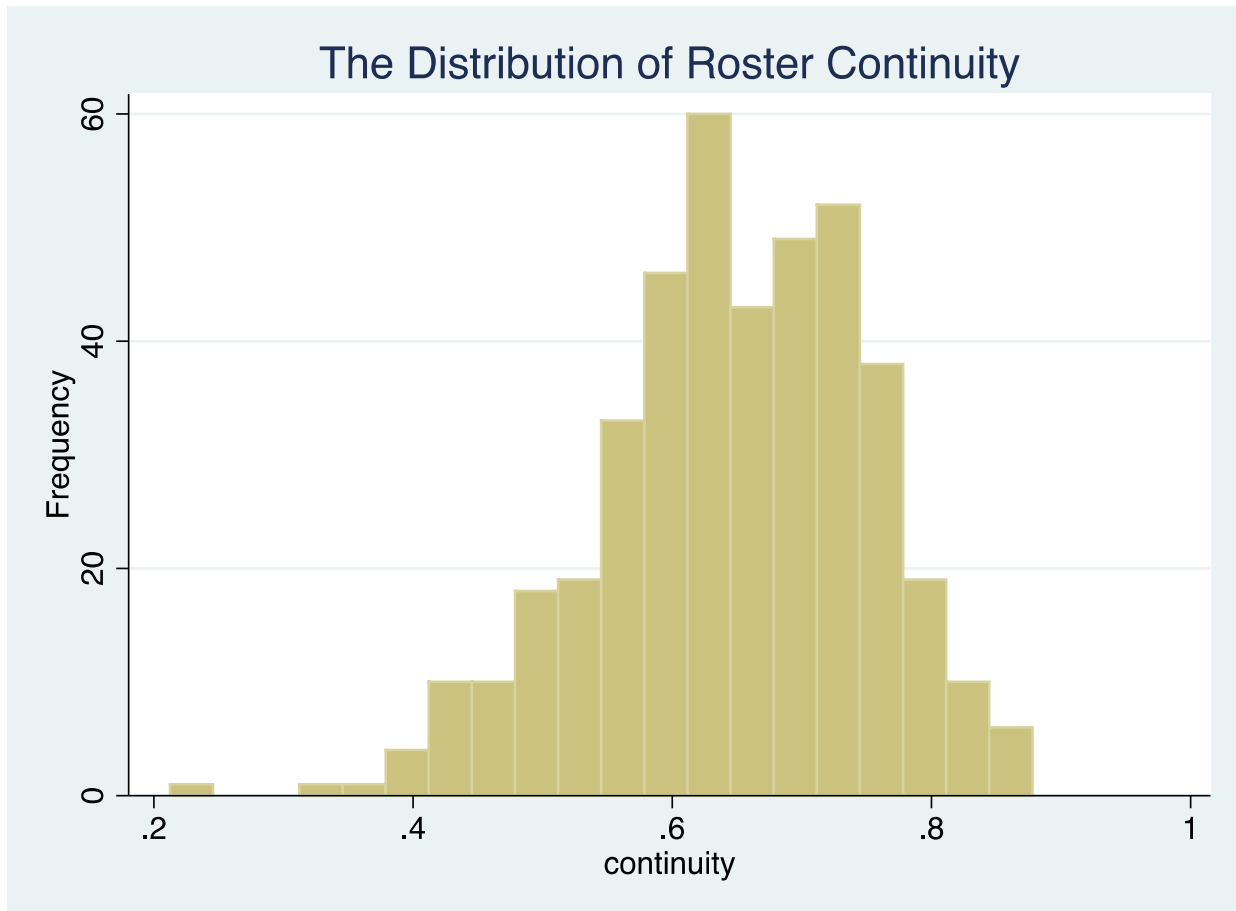
Definition: The proportion of games from the previous season retained by the team.

Units: N/A

Mean=.6473 **Standard Deviation=**.1029

Min=.2125 **25th percentile=**.5843 **Median=**.6521 **75th percentile=**.7264 **Max=**.8779

Histogram:



Variable Name: WL

Original Data Source: Team Standings By Year

Missing Observations: 0/450

Definition: The winning percentage of each team in a season

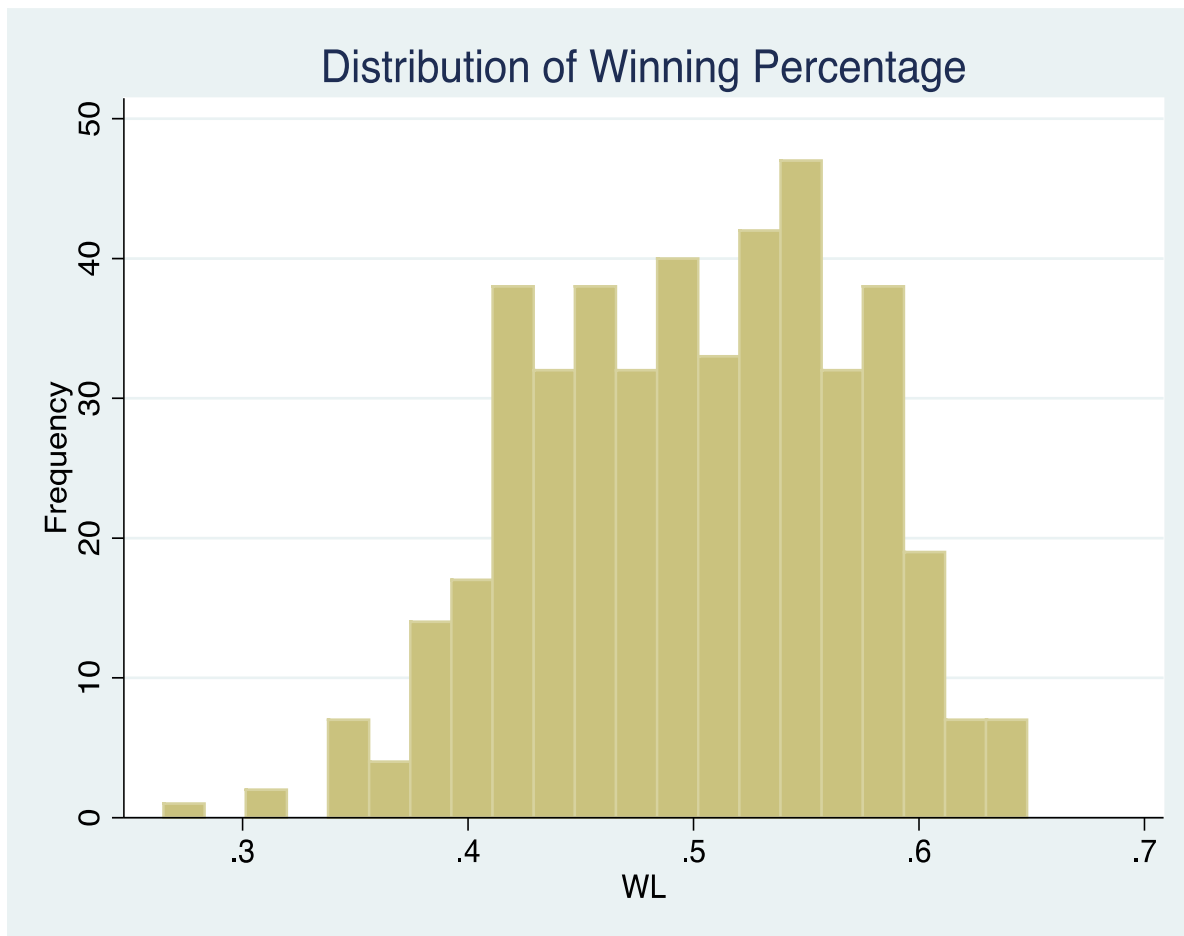
Units: W/(W+L) or % of games won

Mean=.499993

Standard Deviation=.070106

Min=.265 **25th percentile=**.444 **Median=**.5014 **75th percentile=**.556 **Max=**.648

Histogram:



Variable Name: AllStarWL

Original Data Source: Hand Collected Data: All Star W and All Star L variables

Missing Observations: 0/450

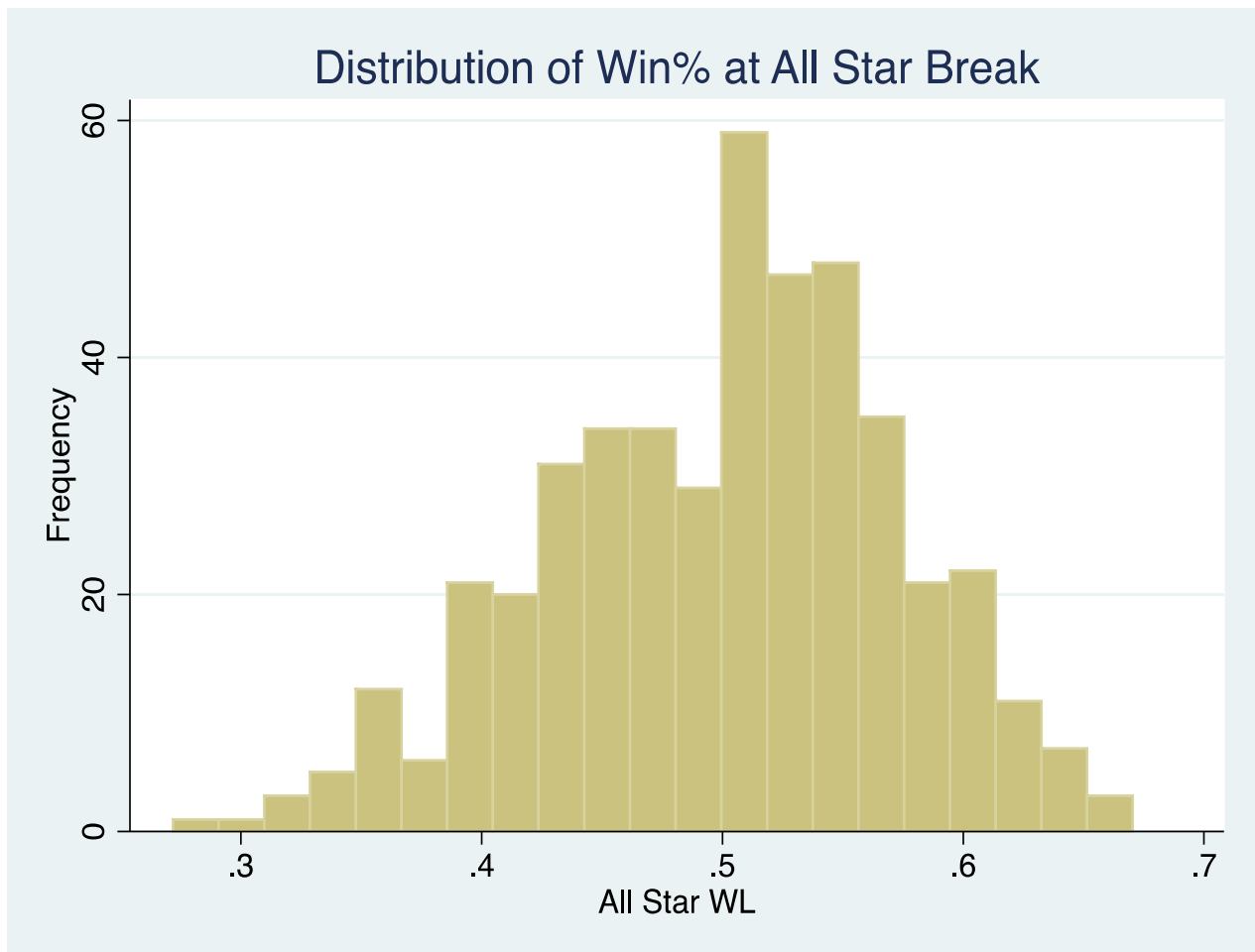
Definition: The winning percentage of each team at the time of the All Star Game

Units: $W/(W+L)$ or % of games won

Mean=.50003 **Standard Deviation**=.072476

Min=.2717 **25th percentile**=.4505 **Median**=.505618 **75th percentile**=.5517 **Max**=.6705

Histogram:



Variable Name: pythWL

Original Data Source: Team Standings By Year

Missing Observations: 0/450

Definition: The Expected winning percentage of each team in a season based on its runs scored and runs allowed in that season.

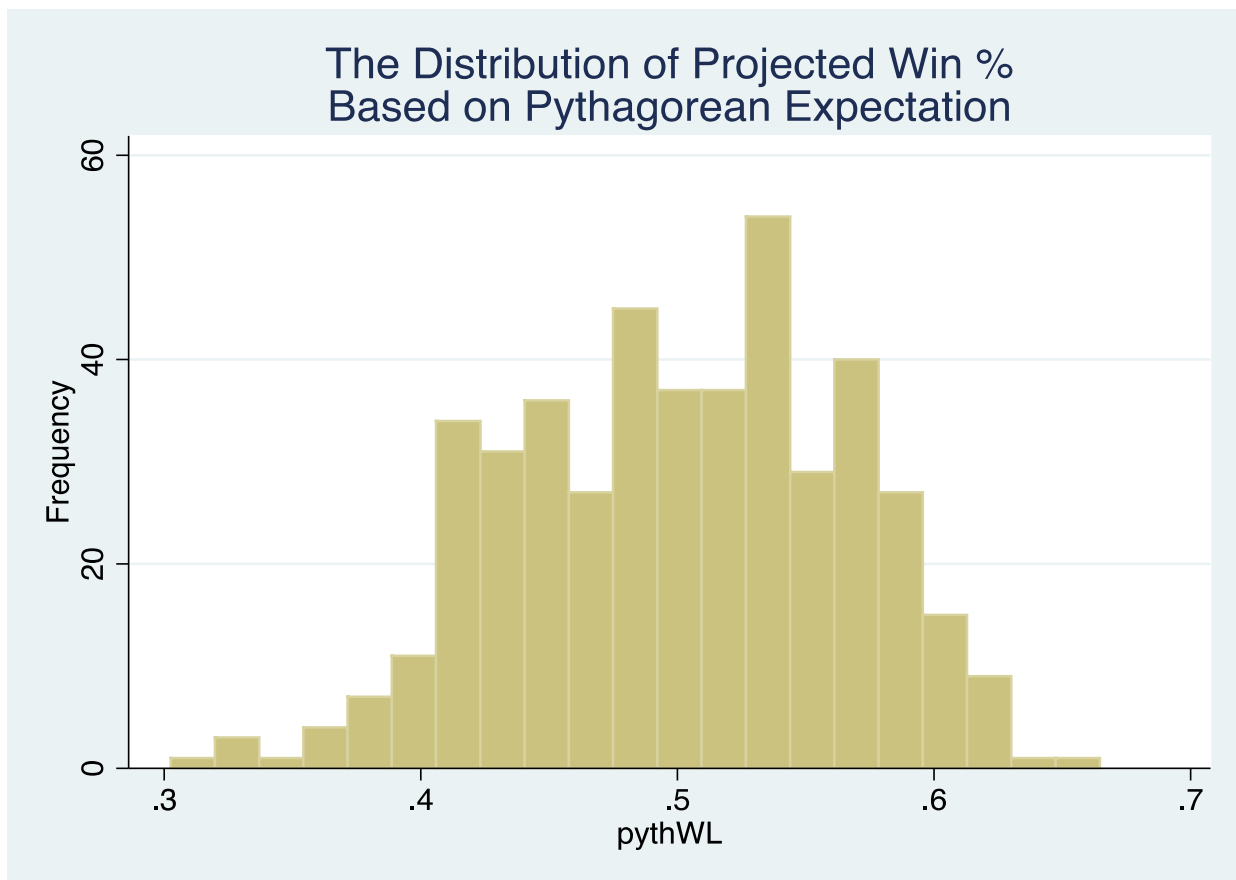
$$pythWL = \frac{(Runs\ Scored)^{1.83}}{(Runs\ Scored)^{1.83} + (Runs\ Allowed)^{1.83}}$$

Units: Percent of games won.

Mean=.500554 **Standard Deviation**=.064563

Min=.302 **25th percentile**=.451 **Median**=.502 **75th percentile**=.549 **Max**=.665

Histogram:



Variable Name: vegasW

Original Data Source: hand collected data

Missing Observations: 60/450

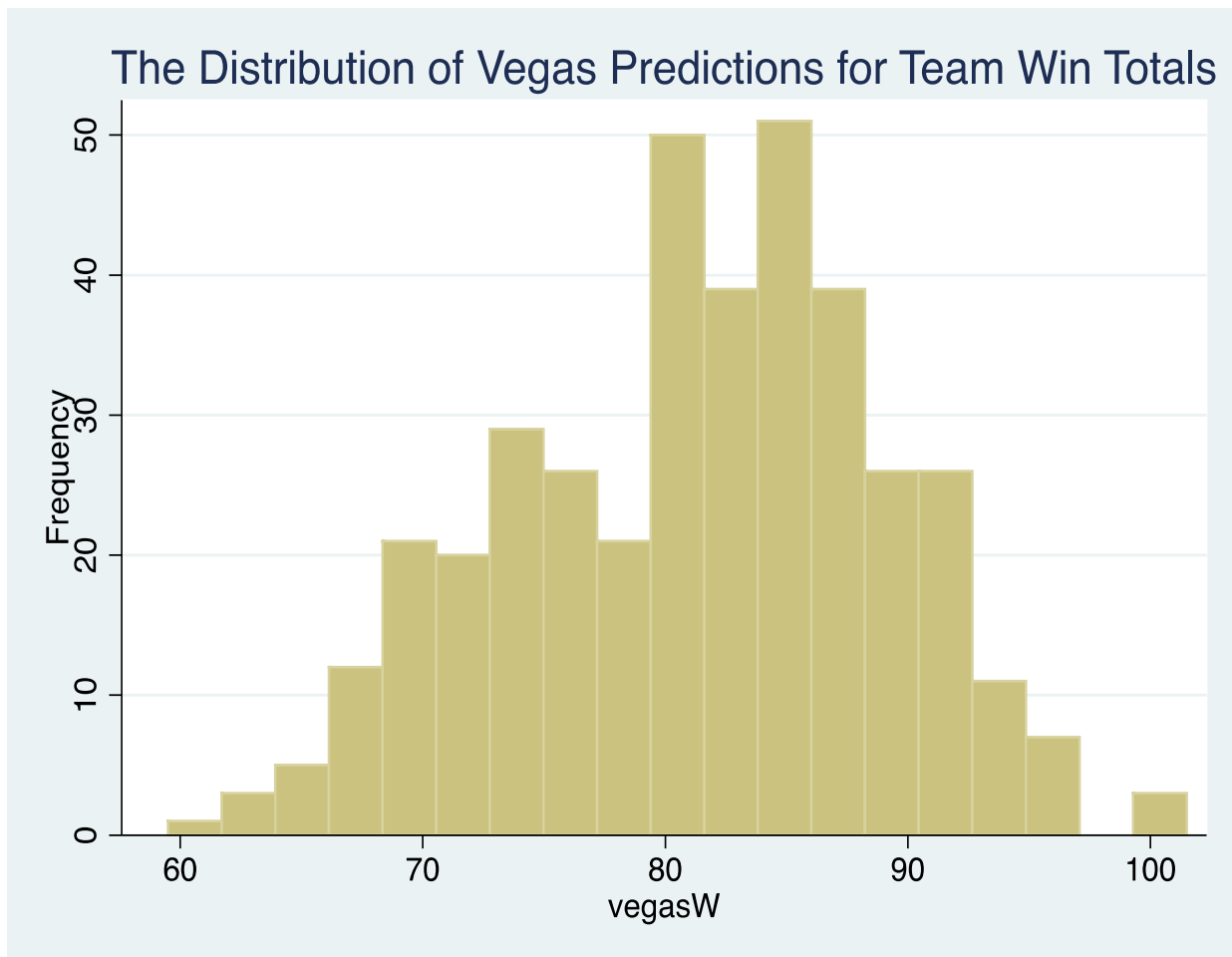
Definition: The over/under win prediction set by casinos in Las Vegas

Units: Games Won

Mean=81.3756 Standard Deviation=7.741

Min=59.5 25th percentile=75.5 Median=82.5 75th percentile=86.5 Max=101.5

Histogram:



Variable Name: W

Original Data Source: Team Standings By Year

Missing Observations: 0/450

Definition: The number of games won by each team in a season

Units: Games Won

Mean=80.9689

Standard Deviation=11.3604

Min=43

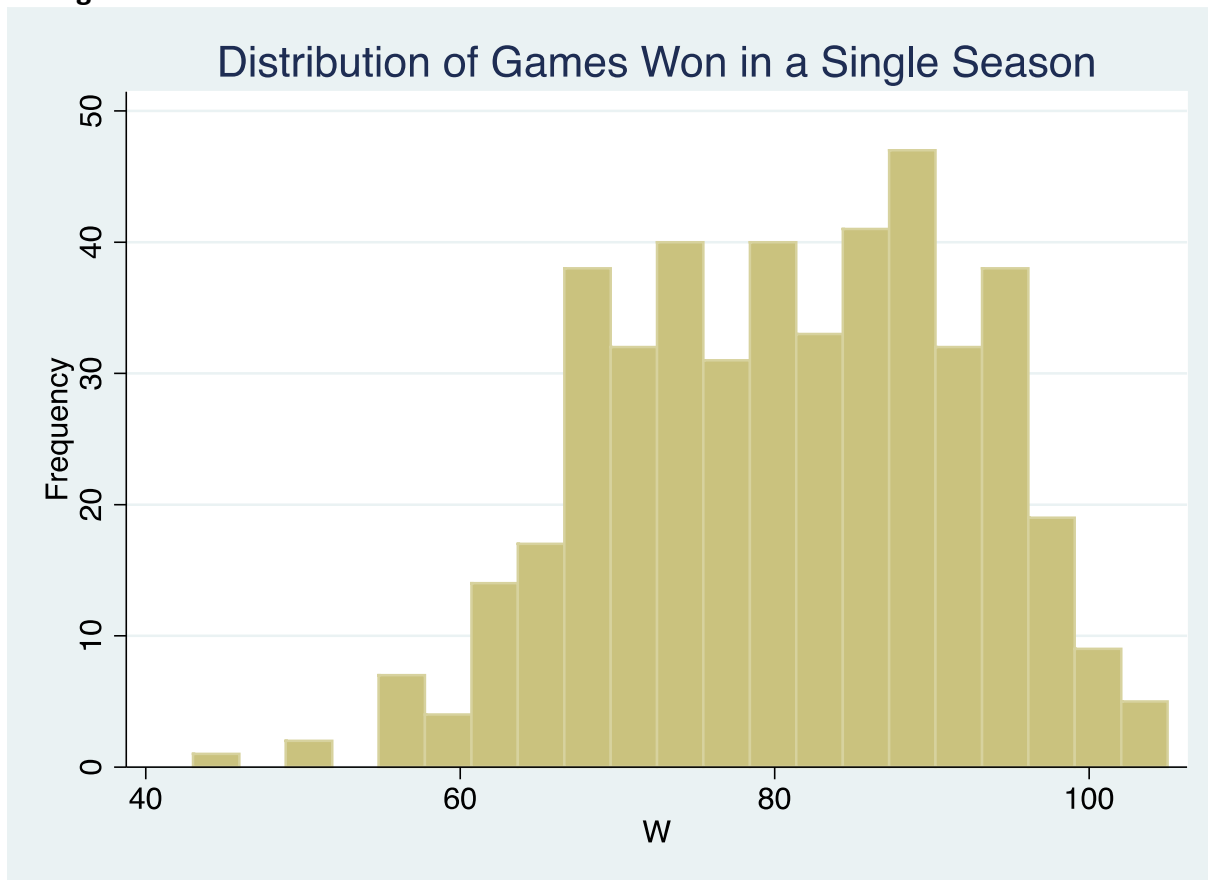
25th percentile=72

Median=81

75th percentile=90

Max=105

Histogram:



Variable Name: L

Original Data Source: Team Standings By Year

Missing Observations: 0/450

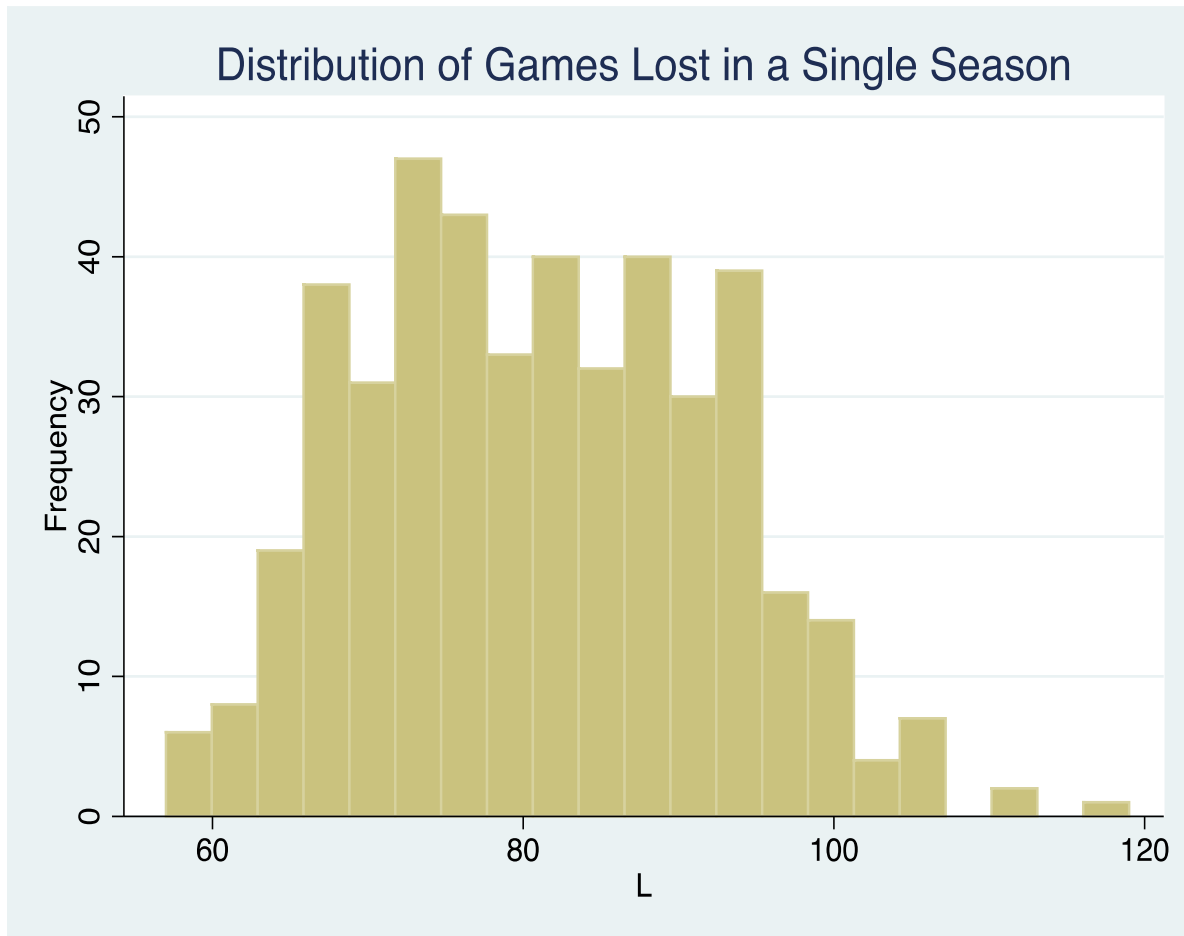
Definition: The number of games lost by each team in a season

Units: Games Lost

Mean=80.9689 **Standard Deviation**=11.3435

Min=57 **25th percentile**=72 **Median**=80.5 **75th percentile**=90 **Max**=119

Histogram:



Variable Name: man_tenure

Original Data Source: Hand Collected Data

Missing Observations: 0/450

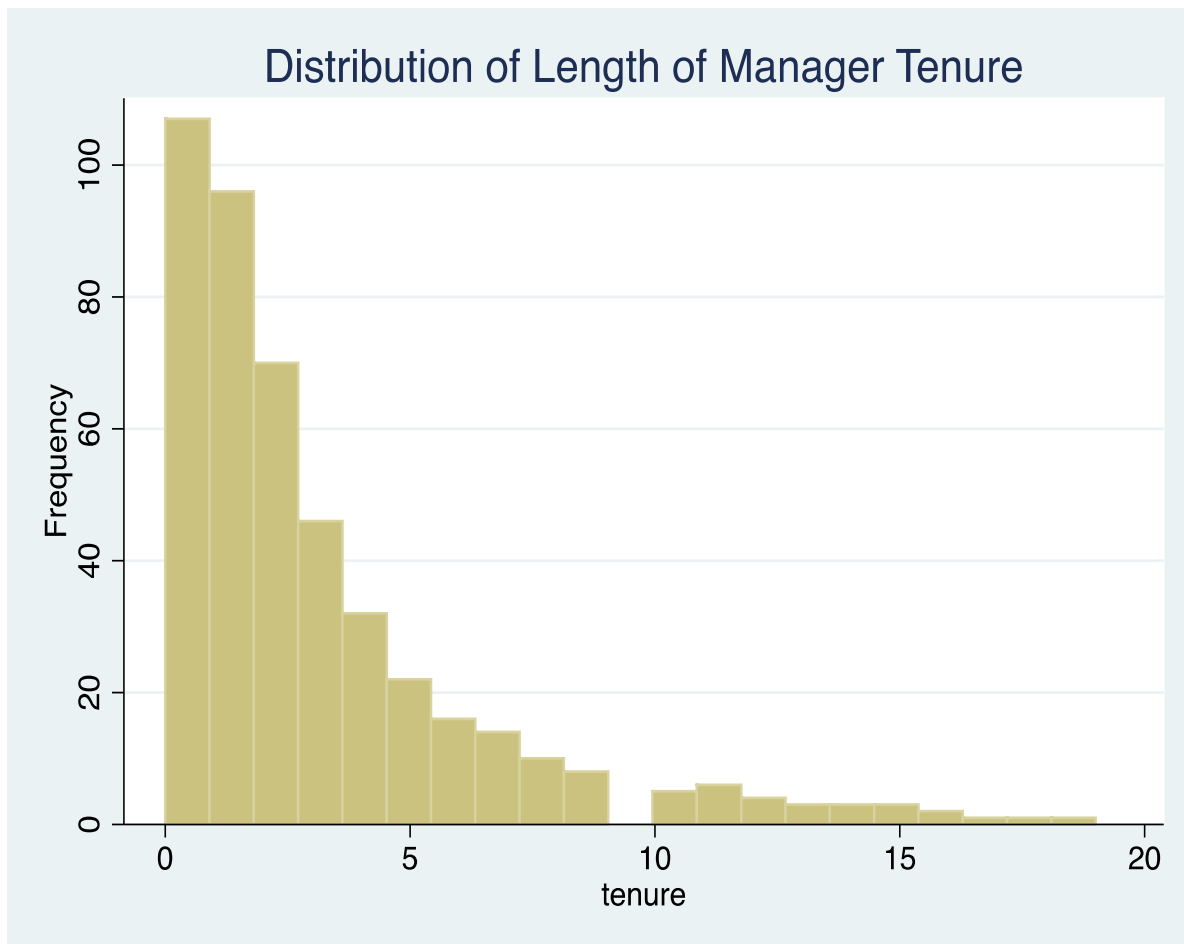
Definition: This variable indicates how many consecutive Opening Days the team's current manager had managed prior to that season with the franchise.

Units: Consecutive years of experience with team

Mean= 2.964 Standard Deviation= 3.50165

Min= 0 25th percentile= 1 Median= 2 75th percentile= 4 Max= 19

Histogram:



Variable Name: OpenWAR

Original Data Source: Player Value by Year Hitter, Player Value by Year Pitcher, Opening Day Rosters

Missing Observations: 30/450

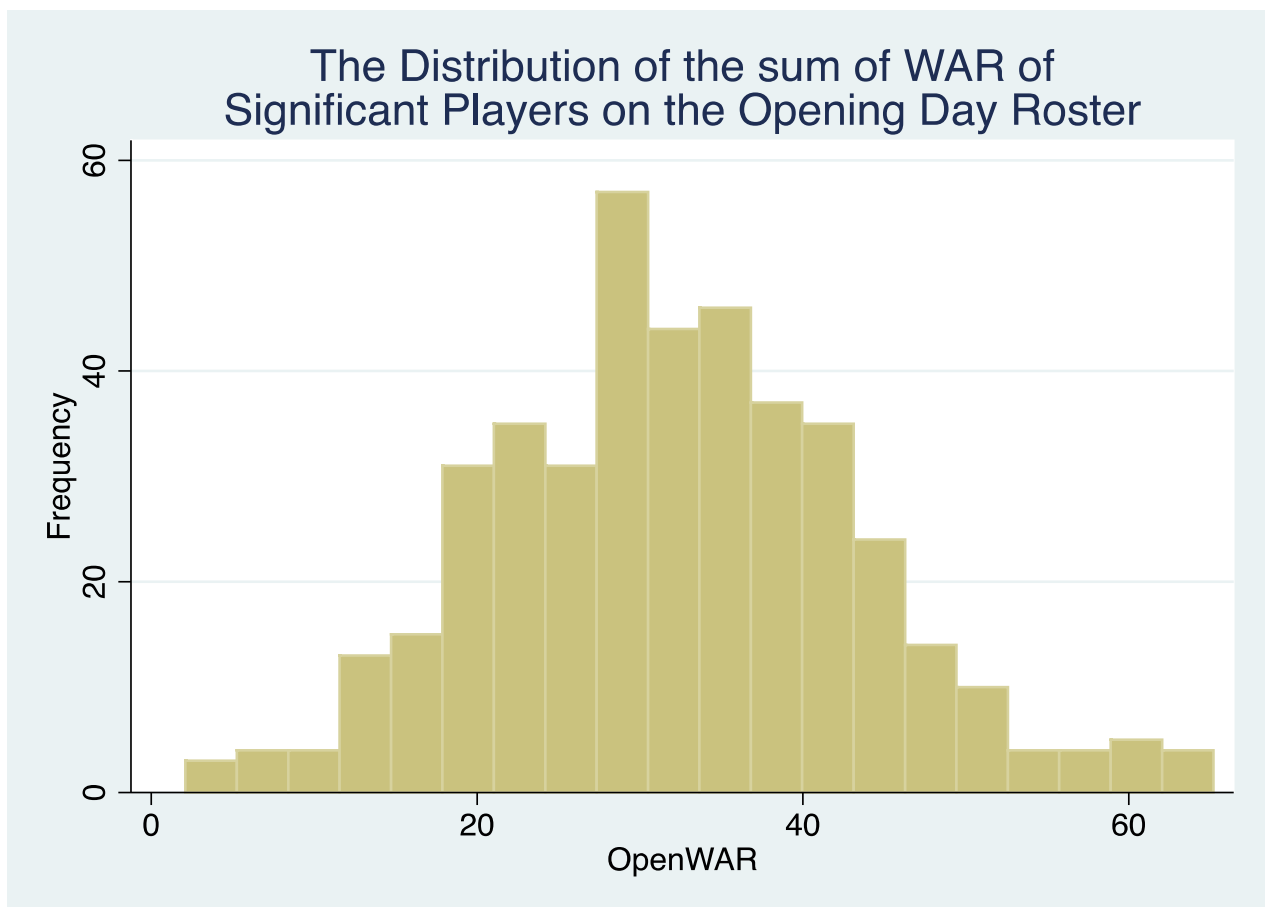
Definition: The sum of the Wins Above Replacement for each significant player on every team's opening day roster for every season. Non-significant players are given a WAR=0.

Units: WAR

Mean= 32.1081 **Standard Deviation=** 11.3218

Min= 2.1 **25th percentile=** 24.25 **Median=** 31.55 **75th percentile=** 39.45 **Max=** 65.2

Histogram:



Variable Name: warexists

Original Data Source: Player Value by Year Hitter, Player Value by Year Pitcher, Opening Day Rosters

Missing Observations: 30/450

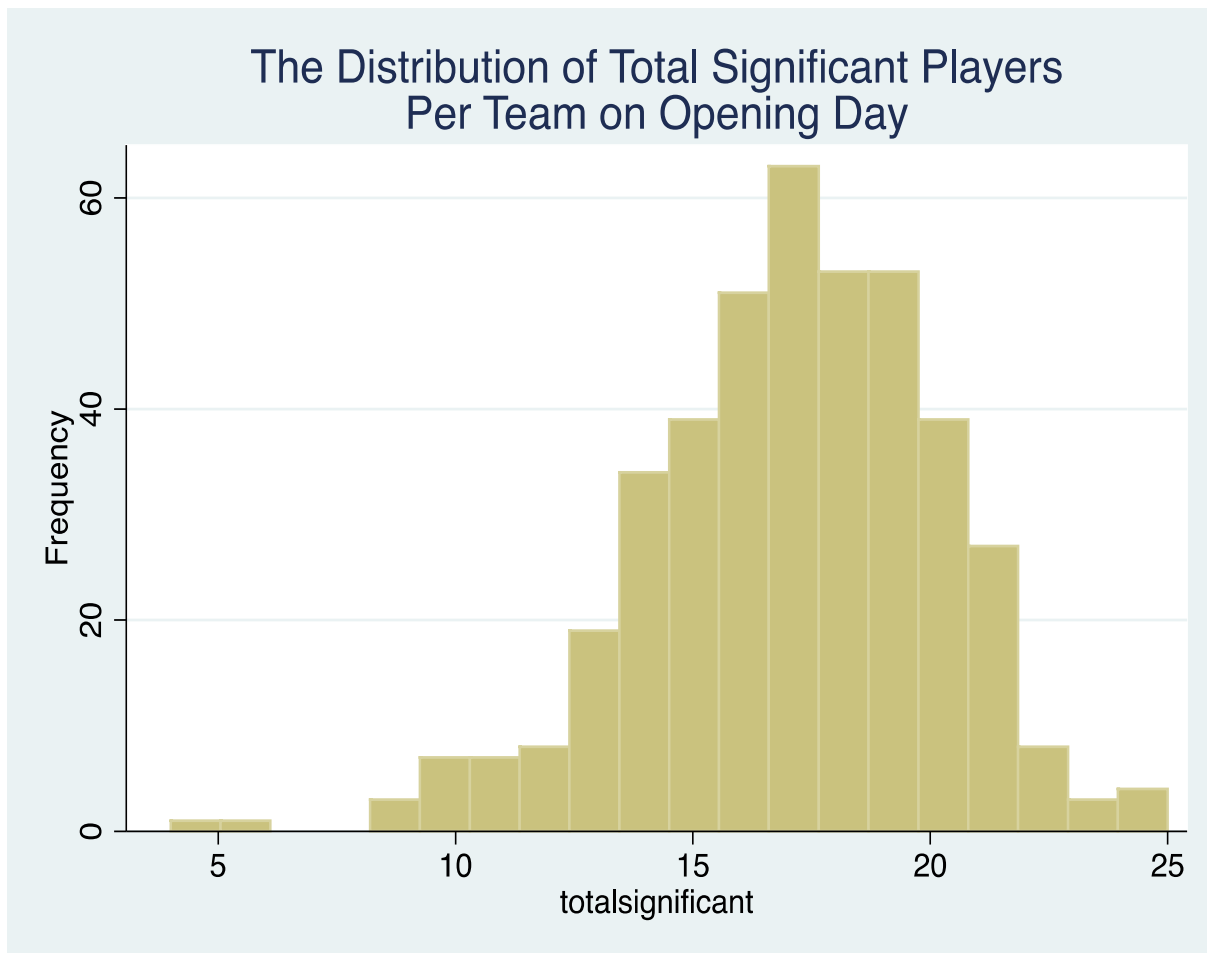
Definition: The number of players on each team's Opening Day roster that were significant in the previous season and therefore have WAR values associated with them.

Units: players

Mean= 16.96 Standard Deviation= 2.96

Min= 4 25th percentile= 15 Median= 17 75th percentile= 19 Max= 25

Histogram:



Variable Name: warpersig

Original Data Source: Player Value by Year Hitter, Player Value by Year Pitcher, Opening Day Rosters

Missing Observations: 30/450

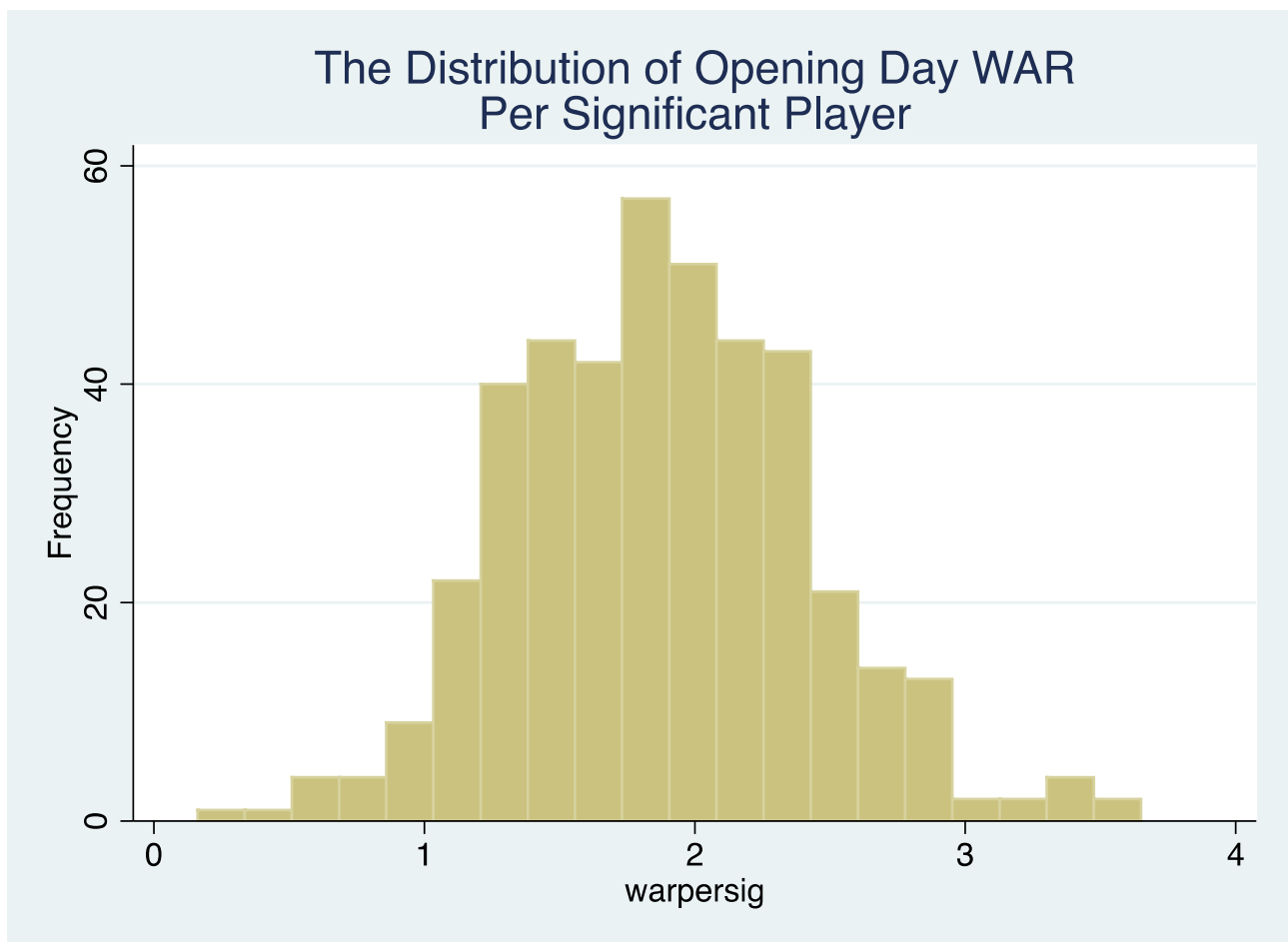
Definition: The Average Opening Day WAR per significant player on the Opening Day Roster.
 $\text{Warpersig} = \text{OpenWAR} / \text{warexists}$

Units: WAR/player

Mean=1.876 Standard Deviation= .548

Min= .162 25th percentile= 1.473 Median= 1.862 75th percentile= 2.239 Max= 3.650

Histogram:



Variable Name: totalplayers

Original Data Source: Opening Day Rosters

Missing Observations: 30/450

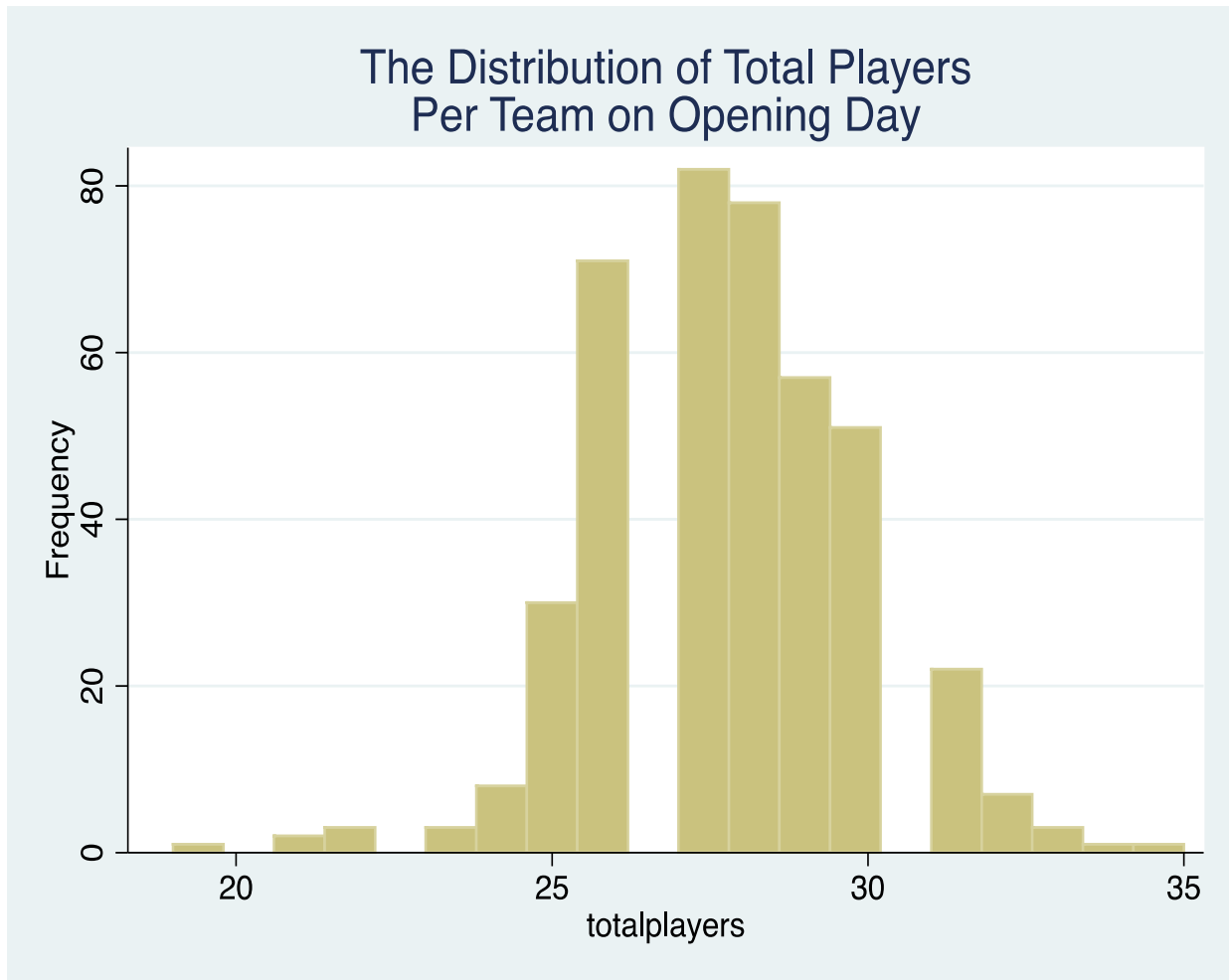
Definition: The total number of players included on recorded as being on each team's Opening Day Roster.

Units: Players

Mean= 27.712 **Standard Deviation=** 2.085

Min= 19 **25th percentile=**26 **Median=** 28 **75th percentile=** 29 **Max=** 35

Histogram:



Variable Name: sigrate

Original Data Source: Player Value by Year Hitter, Player Value by Year Pitcher, Opening Day Rosters

Missing Observations: 30/450

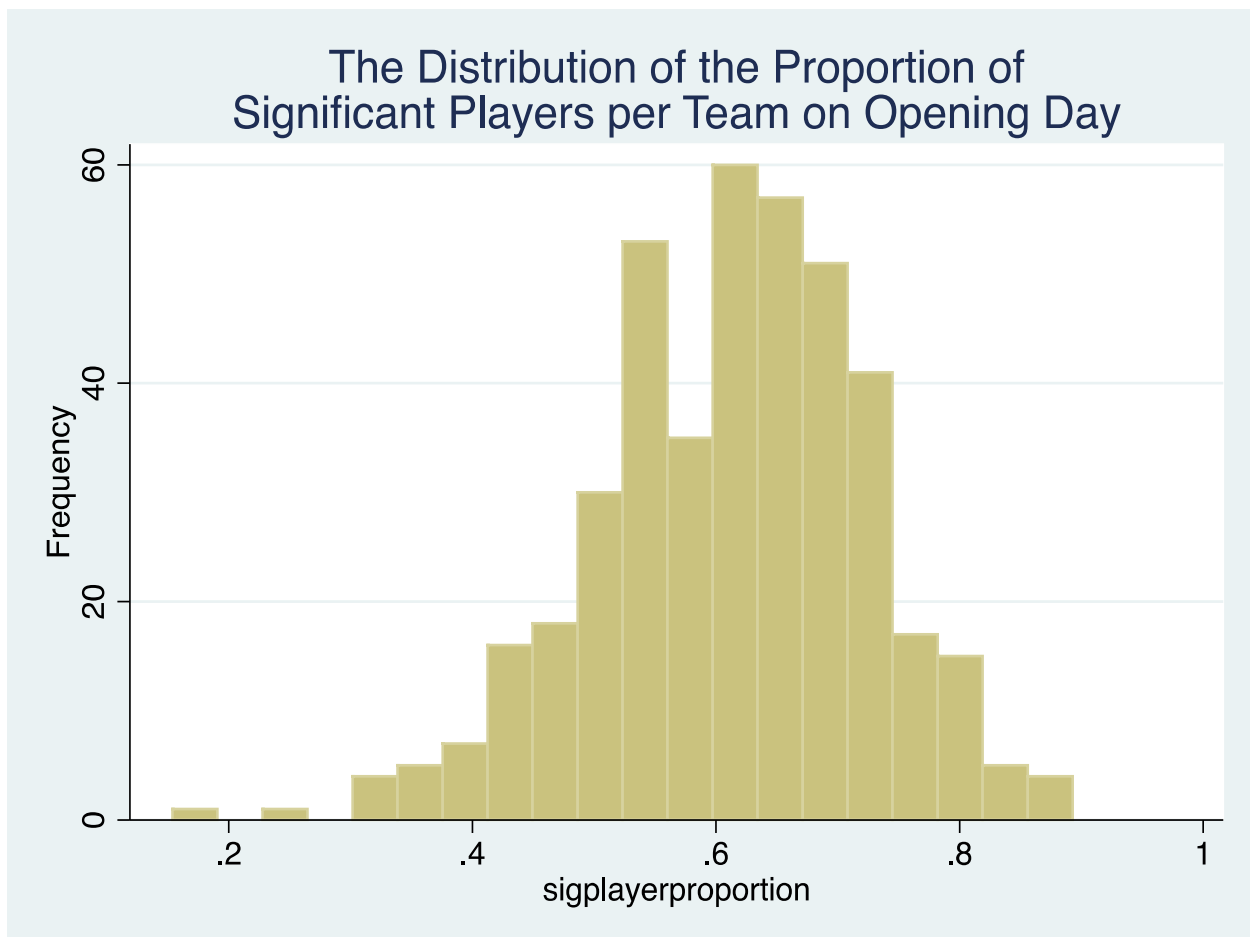
Definition: The proportion of players on each team's Opening Day Roster that were significant in the previous season:
 $\text{Sigrate} = \text{warexists} / \text{totalplayers}$

Units: N/A

Mean=.615 **Standard Deviation=** .112

Min= .154 **25th percentile=** .538 **Median=** .630 **75th percentile=** .692 **Max=** .893

Histogram:



Variable Name: *teampayroll*

Original Data Source: Team Salary

Missing Observations: 0/450

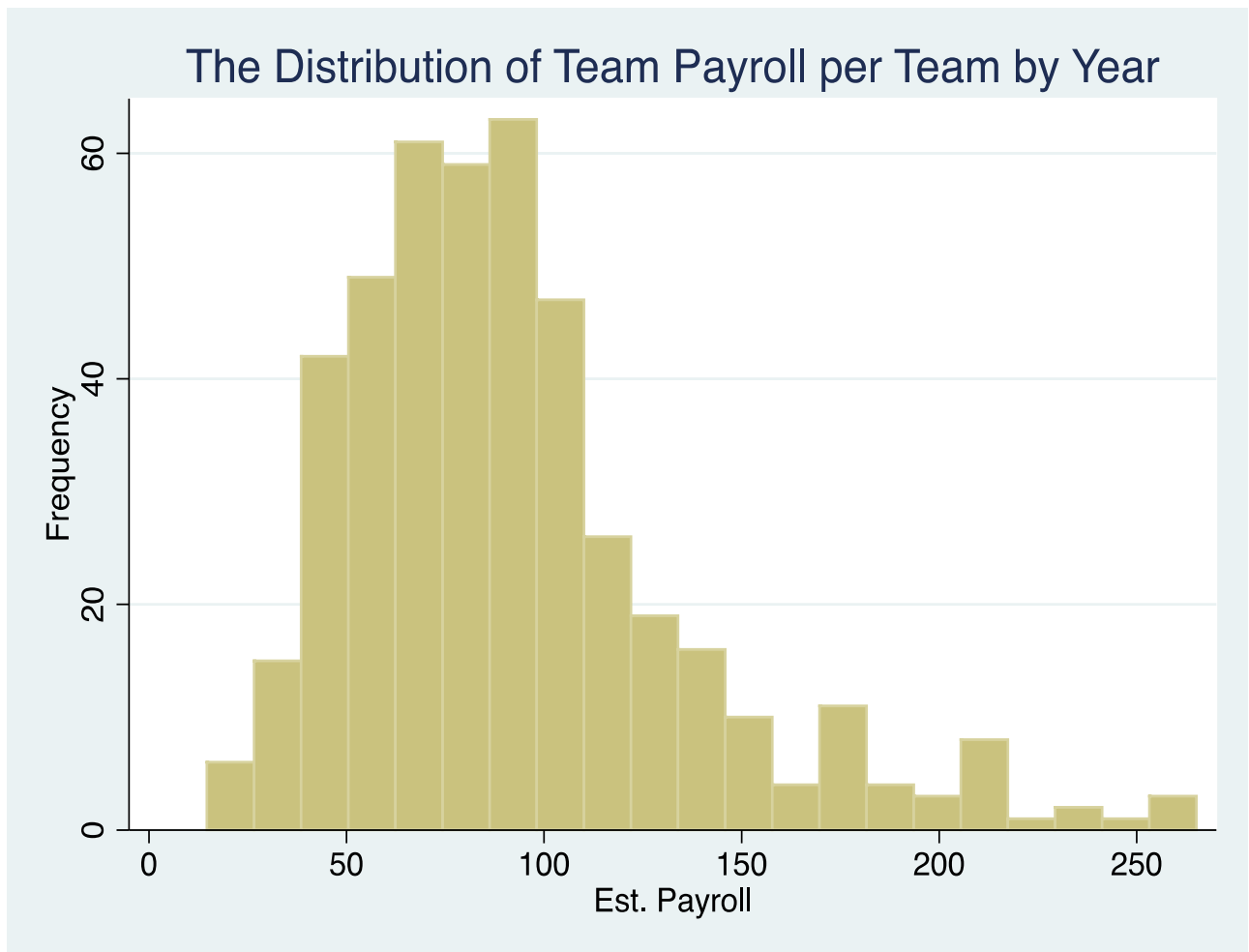
Definition: The estimated amount of money each team paid their players in a given year.

Units: Millions of Dollars

Mean= 91.7918 **Standard Deviation=** 43.03

Min= 14.6715 **25th percentile=** 62.6292 **Median=** 84.832 **75th percentile=** 108.455 **Max=** 265.14

Histogram:



Variable Name: leagueturnover

Original Data Source: Player Value Data by Year Hitters, Player Value Data By Year Pitchers

Missing Observations: 0/450

Definition: This variable indicates how many significant players on a team's Opening Day Roster played for a team from the opposite league in the previous season.

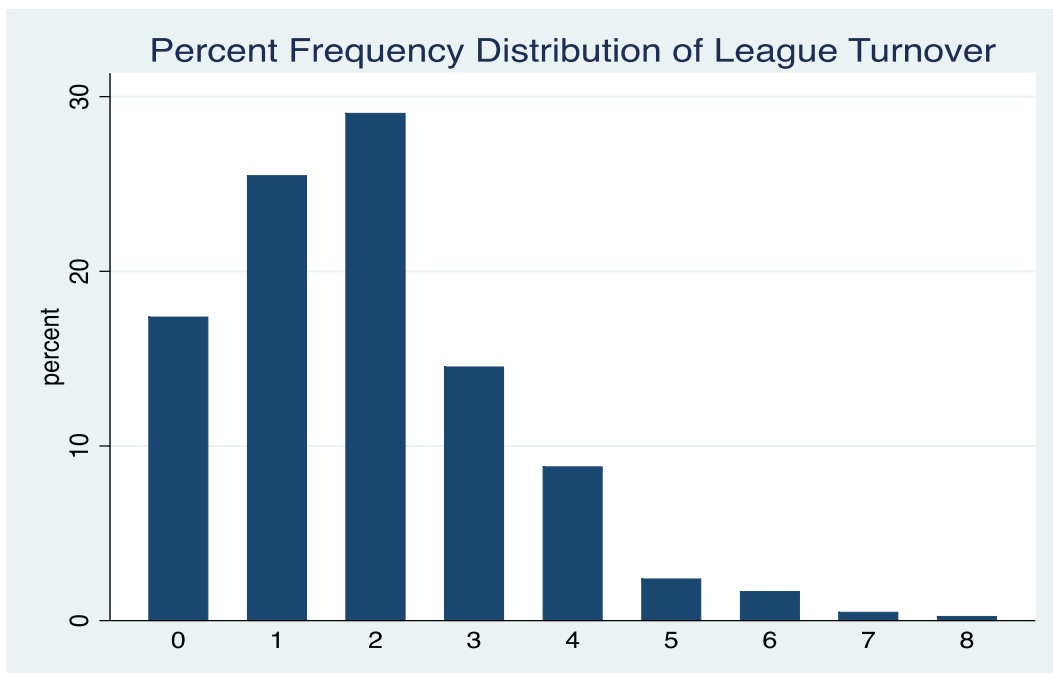
Possible Values: 0,1,2,3,...

Units: Number of Players

Frequency Table

leagueturnover	Freq.	Percent	Cum.
0	73	17.38	17.38
1	107	25.48	42.86
2	122	29.05	71.90
3	61	14.52	86.43
4	37	8.81	95.24
5	10	2.38	97.62
6	7	1.67	99.29
7	2	0.48	99.76
8	1	0.24	100.00
Total	420	100.00	

Bar chart showing percent frequency distribution:



Variable Name: parkfactor

Original Data Source: Team Salary

Missing Observations: 0/450

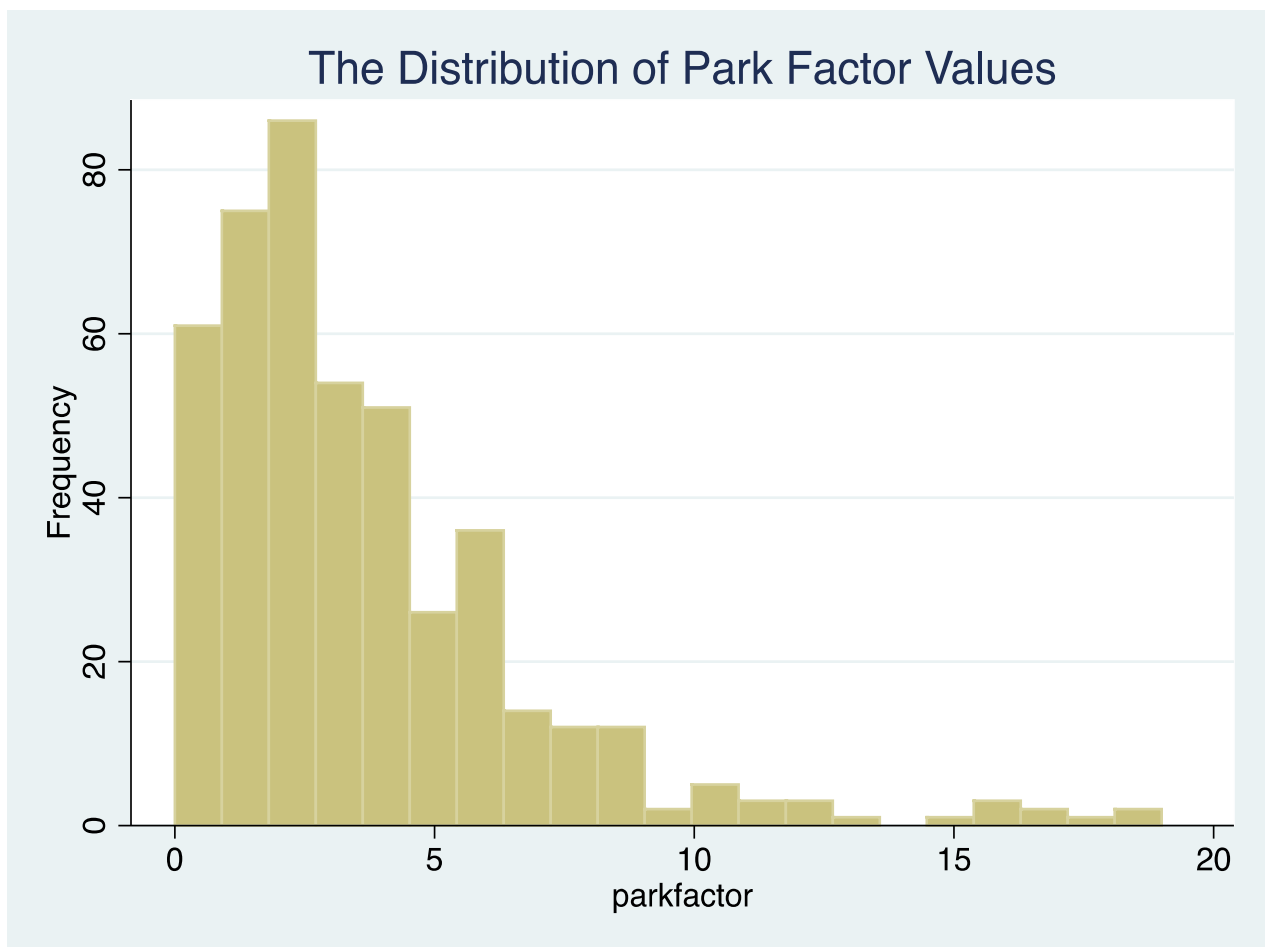
Definition: A measure of the quirkiness of a team's ballpark in terms of how batter or pitcher friendly it is: $\text{parkfactor} = |(ppfactors + bpfactors) / 2 - 100|$

Units: generic

Mean= 3.629 **Standard Deviation=** 3.168

Min=0 **25th percentile=** 1.5 **Median=** 3 **75th percentile=** 5 **Max=** 19

Histogram:



Variable Name: catchcont

Original Data Source: Rosters

Missing Observations: 30/450

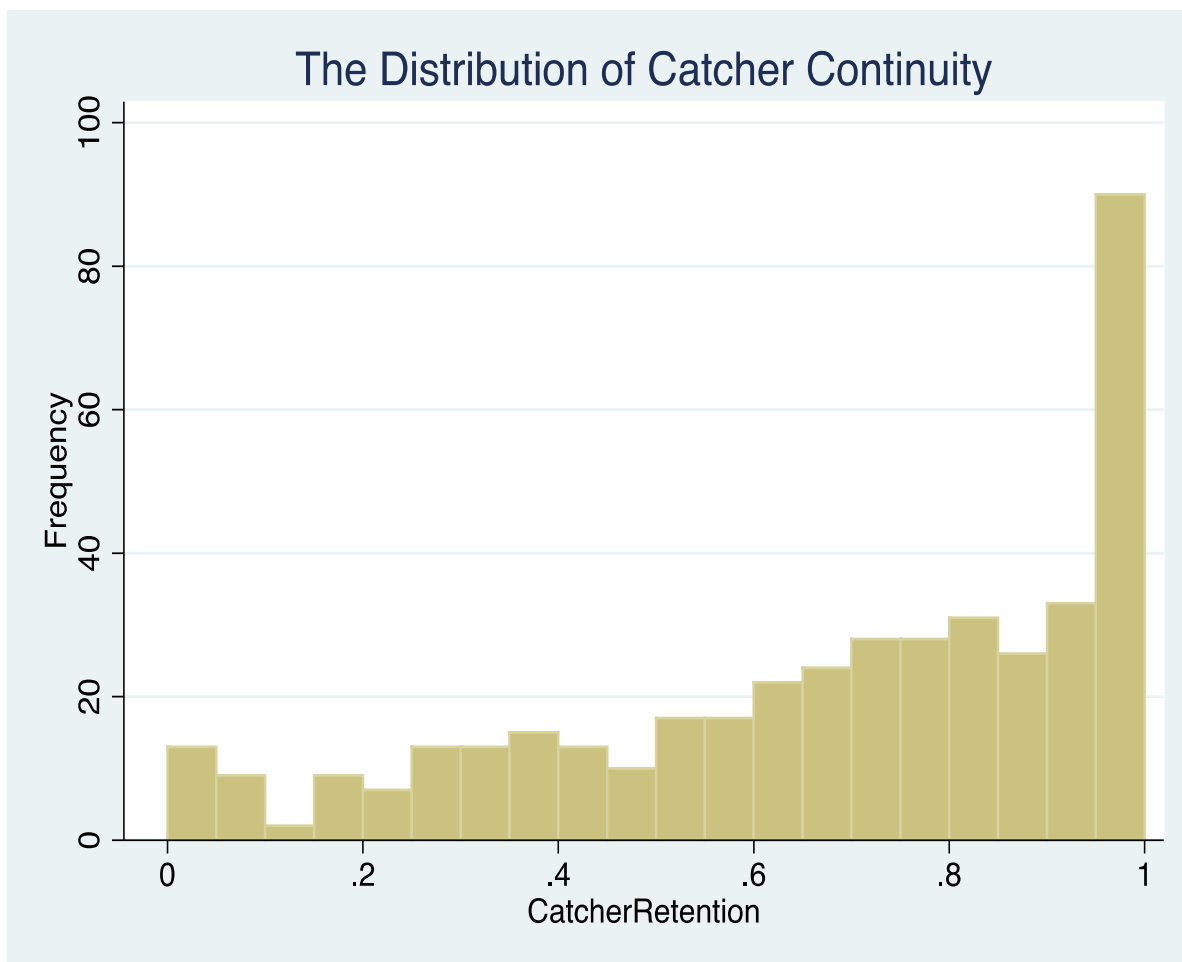
Definition: The proportion of catching appearances from the previous season retained by the team.

Units: N/A

Mean=.6817 **Standard Deviation=**.2823

Min=0 **25th percentile=**.5057 **Median=**.7431 **75th percentile=**.9306 **Max=**1

Histogram:



Variable Name: sscont

Original Data Source: Rosters

Missing Observations: 30/450

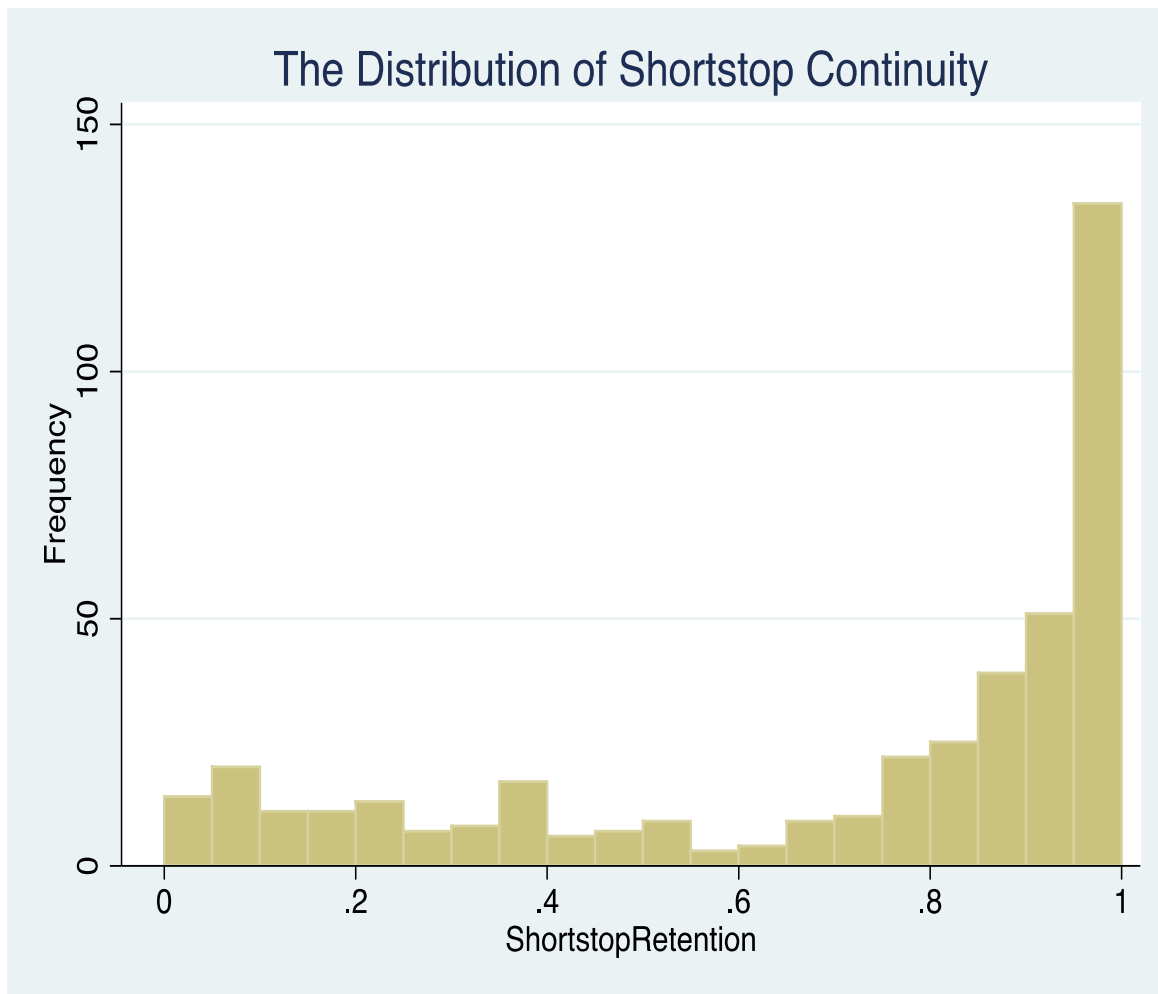
Definition: The proportion of shortstop appearances from the previous season retained by the team.

Units: N/A

Mean=.7111 **Standard Deviation=**.3267

Min=0 **25th percentile=**.4316 **Median=**.8676 **75th percentile=**.9711 **Max=**1

Histogram:



Variable Name: startpitchcont

Original Data Source: Rosters

Missing Observations: 30/450

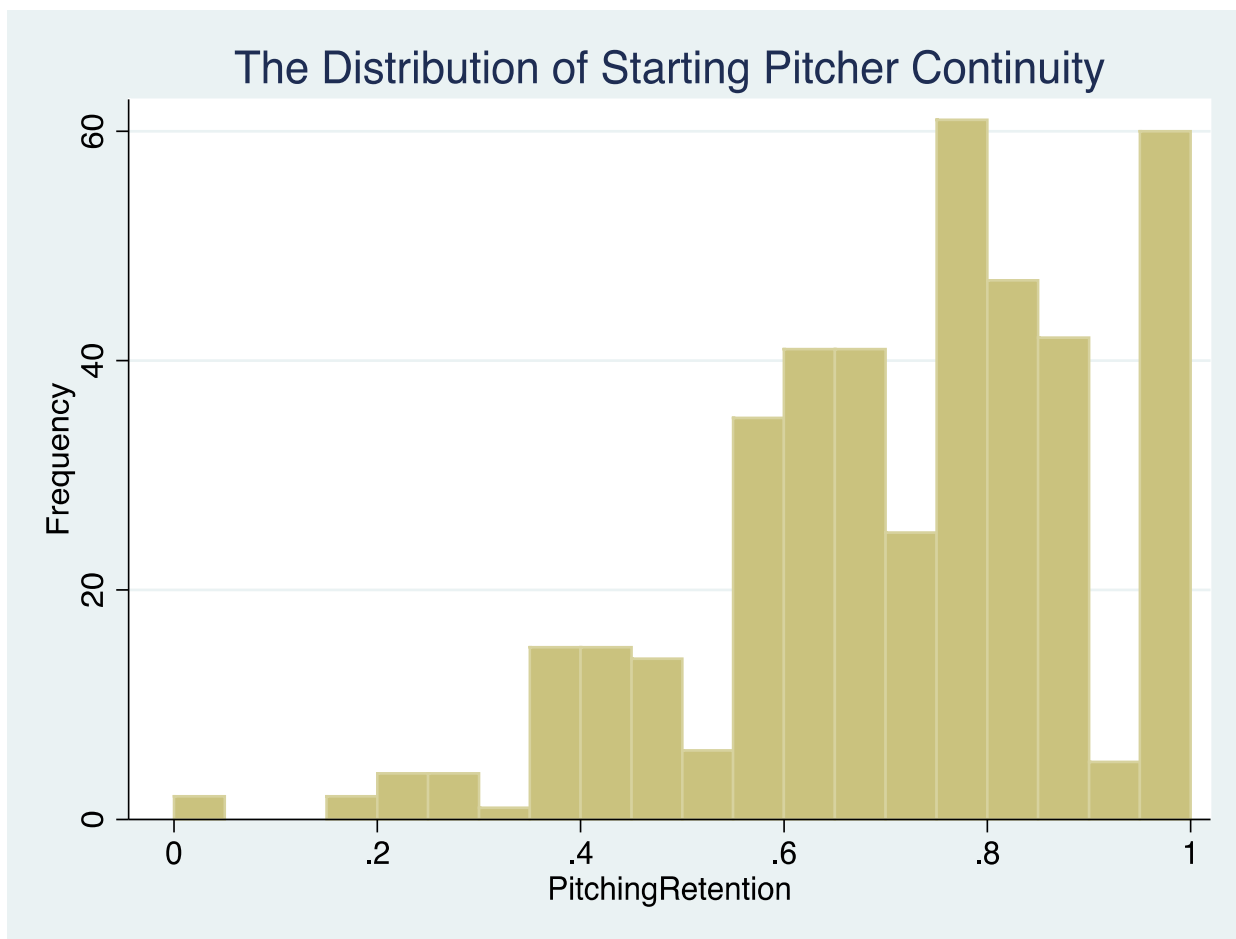
Definition: The proportion of pitching starts by the 5 most frequent starters from the previous season retained by the team.

Units: N/A

Mean=.7229 **Standard Deviation=**.1927

Min=0 **25th percentile=**.6117 **Median=**.7540 **75th percentile=**.8550 **Max=**1

Histogram:



Robustness Appendix

1. OLS Full Season Model without $Turnover * WL_{t-1}$

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 \log(teampayroll)_{t,i} + B_6 WL_{t-1,i} + e_t$$

VIF Test for Multicollinearity:

Variable	VIF	1/VIF
WL		
L1.	2.31	0.433536
warpersig	2.01	0.497162
sigrate	1.49	0.669336
logpayroll	1.45	0.690037
turnover	1.27	0.787931
man_tenure	1.13	0.885983
Mean VIF	1.61	

Breusch-Pagan Test for Heteroskedasticity:

Ho: Constant variance

Variables: fitted values of WL

chi2(1) = **0.45**

Prob > chi2 = **0.5009**

2. OLS Full Season Model with $Turnover * WL_{t-1}$: Turnover and WL_{t-1} are centered

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5 \log(teampayroll)_{t,i} + B_6 WL_{t-1,i} + B_7 Turnover * WL_{t-1} + e_t$$

VIF Test for Multicollinearity

Variable	VIF	1/VIF
cen_WL		
L1.	2.31	0.433329
warpersig	2.01	0.496942
sigrate	1.50	0.668765
logpayroll	1.46	0.687147
cen_turnover	1.28	0.779978
man_tenure	1.13	0.885018
WLint	1.05	0.956805
Mean VIF	1.53	

Breusch-Pagan Test for Heteroskedasticity:

Ho: Constant variance

Variables: fitted values of WL

chi2(1) = **0.31**

Prob > chi2 = **0.5754**

3. OLS Full Season Model with $Turnover * WL_{t-1}$: Turnover and WL_{t-1} are centered

$$AllStarWL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 man_tenure_{t,i} + B_3 warpersig_{t,i} + B_4 sigrate_{t,i} + B_5$$

$$log(teampayroll)_{t,i} + B_6 WL_{t-1,i} + B_7 Turnover * WL_{t-1} + e_t$$

VIF Test for Multicollinearity:

Variable	VIF	1/VIF
cen_WL		
L1.	2.31	0.433329
warpersig	2.01	0.496942
sigrate	1.50	0.668765
logpayroll	1.46	0.687147
cen_turnover	1.28	0.779978
man_tenure	1.13	0.885018
WLint	1.05	0.956805
Mean VIF	1.53	

Breusch-Pagan Test for Heteroskedasticity:

Ho: Constant variance

Variables: fitted values of AllStarWL

chi2(1) = 2.50

Prob > chi2 = 0.1137

4. OLS Full Season Model with Positional Turnover Explanatory Variables

$$WL_{t,i} = B_0 + B_1 catchercont_{t,i} + B_2 sscont_{t,i} + B_3 startpitchcont_{t,i} + B_4 man_tenure_{t,i} + B_5 warpersig_{t,i} + B_6 sigrate_{t,i} + B_7 \log(teampayroll)_{t,i} + B_8 WL_{t-1,i} + e_t$$

VIF Test for Multicollinearity:

Variable	VIF	1/VIF
-----+-----		
WL		
L1.	2.30	0.434837
warpersig	2.02	0.494078
logpayroll	1.45	0.690583
sigrate	1.36	0.736505
man_tenure	1.13	0.888803
startpitchcont	1.11	0.902847
catchcont	1.10	0.907755
sscont	1.07	0.935772
-----+-----		
Mean VIF	1.44	

Breusch-Pagan Test for Heteroskedasticity:

Ho: Constant variance

Variables: fitted values of WL

chi2(1) = 0.41

Prob > chi2 = 0.5237

5. OLS Full Season Model with Park Factor and League Turnover Explanatory Variables:

$$WL_{t,i} = B_0 + B_1 turnover_{t,i} + B_2 parkfactor + B_3(turnover * parkfactor) + B_4 man_tenure_{t,i} + B_5$$

$$warpersig_{t,i} + B_6 sigrate_{t,i} + B_7 log(team payroll)_{t,i} + B_8 WL_{t-1,i} + B_9 league turnover_{t,i} + e_t$$

VIF Test for Multicollinearity:

Variable	VIF	1/VIF
-----+-----		
WL		
L1.	2.36	0.422851
warpersig	2.03	0.493630
sigrate	1.75	0.572818
logpayroll	1.45	0.687806
cen_turnover	1.41	0.710205
leagueturnover	1.20	0.832861
man_tenure	1.13	0.885744
cen_parkfactor	1.02	0.976703
parkint	1.02	0.982239
-----+-----		
Mean VIF	1.49	

Breusch-Pagan Test for Heteroskedasticity:

Ho: Constant variance

Variables: fitted values of WL

chi2(1) = 0.57

Prob > chi2 = 0.4487

XI. REFERENCE LIST

- Becker, Gary. 1975. "Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education." New York: National Bureau of Economic Research.
- Breusch, Trevor, & Pagan, Adrian. (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica: Journal of the Econometric Society*, 1287-1294.
- Davenport, Clay, & Woolner, Keith. 1999. "Revisiting the Pythagorean Theorem." *Baseball Prospectus*, June 30. <http://www.baseballprospectus.com/article.php?articleid=342>
- Druschel, Henry. 2016. "A guide to the Projection Systems." Retrieved March 30, 2017, from <http://www.beyondtheboxscore.com/2016/2/22/11079186/projections-marcel-pecota-zips-steamer-explained-guide-math-is-fun>
- FanGraphs. (n.d.). "Projection Systems." Retrieved March 30, 2017, from <http://www.fangraphs.com/library/principles/projections/>
- Glebbeek, Arie, and Erik Bax. 2004. "Is High Employee Turnover Really Harmful? An Empirical Testing Using Company Records." *Academy of Management Journal*, 47(2): 277-286.
- Glenn, Andrew, Joseph P. McGarrity, and Jim Weller. 2001. "Firm-specific Human Capital, Job Matching, and Turnover: Evidence from Major League Baseball, 1900-1992." *Journal of Economic Inquiry*, 39(1): 86-93.
- Harris, Mark N., Kam Ki Tang, and Yi-Ping Tseng. 2006. "Employee Turnover: Less is not Necessarily more?" *Contributions to Economic Analysis*, 274: 327-349.
- Hill, Gregory C. 2009. "The Effect of Frequent Managerial Turnover on Organizational

- Performance: A Study of Professional Baseball Managers.” *The Social Science Journal*, 46(3): 557-570.
- Jovanovic, Boyan. 1979. “Job Matching and the Theory of Turnover.” *Journal of Political Economy*, 87(5): 972–990.
- Kahane, Leo, and Stephen Shmanske. 1997. “Team Roster Turnover and Attendance in Major League Baseball.” *Applied Economics*, 29(4): 425-431.
- Land of Basketball. (n.d.). “NBA Teams: Cleveland Cavaliers Records Year by Year.” Retrieved April 03, 2017, from http://www.landofbasketball.com/cleveland_cavaliers
- Major League Baseball. (n,d.). “Glossary.” Retrieved March 30, 2017, from <http://m.mlb.com/glossary/projection-systems>
- Morse, Alan L., Stephen L. Shapiro, Chad D. McEvoy, and Daniel A. Rascher. 2008. “The Effects of Roster Turnover on Demand in the National Basketball Association.” *International Journal of Sport Finance*, 3(1): 8-18.
- Robinson, Cecil, and Randall E. Schumacker. 2009. "Interaction effects: centering, variance inflation factor, and interpretation issues." *Multiple Linear Regression Viewpoints*, 35(1): 6-11.
- Siebert, Stanley W., and Nikolay Zubanov. 2009. “Searching for the Optimal Level of Employee Turnover: A Study of a Large U.K. Retail Organization.” *Academy of Management Journal*, 52(2): 294-313.
- Sports Reference LLC. (n,d.). “Pythagorean Theorem.” *Baseball-Reference.com - Major League Statistics and Information*. Retrieved March 31, 2017, from http://www.baseball-reference.com/bullpen/Pythagorean_Theorem_of_Baseball

XII. DATA SOURCES

Sports Odds History. Retrieved from <http://www.sportsoddshistory.com/mlb-regular-season-win-total-results-by-team/>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2016-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2016-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2015-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2015-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2014-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2014-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2013-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2013-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2012-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2012-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2011-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2011-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2010-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2010-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2009-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2009-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2008-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2008-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2007-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2007-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2006-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2006-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2005-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2005-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2004-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2004-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2003-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2003-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2002-misc_teams_miscellaneous.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2002-misc.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2016-standings_expanded_standings_overall.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2016-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2015-standings_expanded_standings_overall.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2015-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2014-standings_expanded_standings_overall.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2014-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2013-standings_expanded_standings_overall.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2013-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2012-standings_expanded_standings_overall.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2012-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2011-standings_expanded_standings_overall.csv] Retrieved from

<http://www.baseball-reference.com/leagues/MLB/2011-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2010-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2010-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2009-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2009-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2008-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2008-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2007-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2007-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2006-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2006-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2005-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2005-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2004-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2004-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2003-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2003-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2002-standings_expanded_standings_overall.csv] Retrieved from
<http://www.baseball-reference.com/leagues/MLB/2002-standings.shtml>

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2016-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2016-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2015-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2015-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2014-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2014-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2013-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2013-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2012-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2012-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2011-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2011-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2010-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2010-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2009-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2009-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2008-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2008-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2007-value-batting_players_value_batting.csv] Retrieved from

http://www.baseball-reference.com/leagues/MLB/2007-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2006-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2006-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2005-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2005-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2004-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2004-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2003-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2003-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2002-value-batting_players_value_batting.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2002-value-batting.shtml#players_value_batting::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2016-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2016-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2015-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2015-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2014-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2014-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2013-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2013-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2012-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2012-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2011-value-pitching_players_value_pitching.csv] Retrieved from

http://www.baseball-reference.com/leagues/MLB/2011-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2010-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2010-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2009-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2009-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2008-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2008-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2007-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2007-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2006-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2006-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2005-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2005-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2004-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2004-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2003-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2003-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[leagues_MLB_2002-value-pitching_players_value_pitching.csv] Retrieved from
http://www.baseball-reference.com/leagues/MLB/2002-value-pitching.shtml#players_value_pitching::none

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

[teams_[TEAM]_[YEAR]-roster_appearances.csv] Retrieved from
[http://www.baseball-reference.com/teams/\[TEAM\]/\[YEAR\]-roster.shtml](http://www.baseball-reference.com/teams/[TEAM]/[YEAR]-roster.shtml)

Sports Reference LLC. *Baseball-Reference.com - Major League Statistics and Information.*

<http://www.baseball-reference.com/>.

USA Today. (2016). MLB Salary. Retrieved from

<http://www.usatoday.com/sports/mlb/salaries/2016/player/all/>