**Does Option Theory Hold
In Major League Baseball Contracts?**

by

Alexander Gross

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Honors Bachelor of Sciences in Mathematics and Economics with Distinction

Spring 2014

© 2014 Alexander Gross
All Rights Reserved

**Does Option Theory Hold
In Major League Baseball Contracts?**

by

Alexander Gross

Approved: _____
Dr. Charles Link, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Dr. Daniel Brown, Ph.D.
Committee member from the Department of Public Policy and
Administration

Approved: _____
Dr. Meryl Gardner, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved: _____
Michael Arnold, Ph.D.
Directory, University Honors Program

ACKNOWLEDGMENTS

I would first like to thank Dr. Charles Link, who advised me through this whole process and graciously allowed me to use his dataset in this endeavor. He was excited to work with me from day one, and I thank him for giving me the opportunity to write this paper.

I also thank Dr. Daniel Brown for all of his technical guidance. His help in defining variables and performing data transformations in STATA proved most useful. He was always available to discuss the particulars of my thesis.

Finally, I would like to thank my family and friends. They have provided nothing but support the past four years and I dedicate this paper to them as a culmination of my studies at Delaware.

TABLE OF CONTENTS

LIST OF TABLES	v
ABSTRACT	vi
1 INTRODUCTION	1
2 BASIC OPTION THEORY, MLB CONTRACTS, AND THE LABOR MARKET	3
2.1 Basic Option Theory.....	3
2.2 MLB Contracts and the Labor Market	5
3 LITERATURE	9
4 DATA AND METHODOLOGY	11
4.1 Data.....	11
4.2 Methodology.....	15
4.2.1 Option Buying Model.....	16
4.2.2 Option Exercising Model	17
5 RESULTS.....	21
5.1 Option Buying Model.....	21
5.2 Option Exercising Model	27
6 CONCLUSION	33
REFERENCES	35
A WAGE REGRESSION FOR EXERCISING MODEL.....	36

LIST OF TABLES

Table 1: Option Frequencies for Free Agent Hitters, 2003-2011	8
Table 2: Variable Summary Statistics for Option Buying Model	17
Table 3: Cost-Benefit Analysis of Exercising Option with Buyout	19
Table 4: Variable Summary Statistics for Option Exercising Model	20
Table 5: Probability of Buying Club or Vesting Option on Player's Services	21
Table 6: Marginal Effects of <i>STDV_OPS_LAG</i> on the Probability of Buying a Club or Vesting Option	24
Table 7: Probability of Exercising a Club or Vesting Option on Player's Services	27
Table 8: Marginal Effects of <i>logRealSalaryOpDiff</i> on the Probability of Exercising a Club or Vesting Option	30
Table 9: Effect of Performance Variables on Log Real Salary	36

ABSTRACT

Major League Baseball teams use contract options in order to obtain the right to buy a player's services for an extra year. Contract options seem to be theoretically similar to financial options, which reduce the risk of buying volatile assets. In this paper, I show that teams follow option theory as defined by financial options to an extent, both in terms of buying and exercising options. The sample of players consists of free agent hitters between the years 2003 and 2011. I use two separate probit regressions to model the option process. First, I estimate the probability that a new contract has a club or vesting option, with volatility being measured by the standard deviation of monthly on base plus slugging percentage for the previous season. I then estimate the probability that an option is exercised, using the difference between the real predicted salary and real option salary minus real buyout as an indicator of return to performance above the strike price. I find that the coefficient of standard deviation of monthly OPS is positive and significant at the 5% level and that the coefficient of log real salary-option difference is positive and significant at the 5% level, which shows that Major League Baseball teams do treat contract options like financial options to an extent. I conclude that teams could use options more efficiently by making player volatility a more impactful factor when deciding to buy options.

Chapter 1

INTRODUCTION

Major League Baseball (MLB) has been a fertile field for economic research in the past due to its organized labor market and standardization of play, which leads to repeated experiments. MLB contracts specify the salary paid to a player for any given number of years. Theoretically, a player's salary is supposed to reflect that player's value to the team, just as a worker is paid in accordance to the value added to a business. However, determining the true value of a player is an uncertain process, especially when trying to predict performance several years into the future. Therefore, teams have introduced various contract clauses throughout the years to reduce the risk of uncertainty. For example, performance incentives provide bonuses to players for reaching certain goals, but often at the expense of guaranteed salary.

Another risk-reducing contract element is an option. Options are best known from their application in finance, where an option is defined as the *right* to buy the underlying asset at a certain price at a certain time in the future (a "call option"). Options are seen as a risk-reduction tool because the buyer has the right, but not the necessity, of buying the underlying asset at the specified time. Therefore, if the asset falls sharply in value, the option buyer can choose to not exercise the option (i.e., not buy the underlying asset at the given price) and lose only the cost of the option. This fact makes options particularly attractive when the underlying asset is volatile (high risk/high reward).

MLB contract options appear to work similarly: a team can buy the right to buy a player's services for another year, where the player's services are seen as an underlying asset. However, do teams treat contract options as an investor treats a stock option? That is, do teams take volatility into account when deciding to buy an option, and do they compare the value of the player to the option value when exercising the option? To model this problem, I use two probit regressions to estimate the two segments of the option decision: whether to buy an option to extend the contract by a year and whether to exercise an option. This question is relevant because if teams must pay players to buy an option on their services, then teams will want to be as efficient as possible (i.e., conform to financial option theory) in using options. This study attempts to show whether or not teams are efficient in both buying and exercising options, which is important in regard to teams' budget constraints. In addition, many studies have been done on the returns to performance in different settings, but only a few studies have even touched upon the use of options in MLB contracts. This topic therefore enters uncharted research territory.

The rest of the paper is as follows. Chapter 2 discusses option theory and MLB contracts in greater detail. Chapter 3 is a literature review, outlining the studies mentioned above as well as other relevant studies. Chapter 4 is a description of the data, sample selection, and methodology. Chapter 5 contains the results of the regressions. Chapter 6 concludes.

Chapter 2

BASIC OPTION THEORY, MLB CONTRACTS, AND THE LABOR MARKET

2.1 Basic Option Theory

To understand how MLB teams use contract options, one must first understand the basics of option theory. I draw my definitions from a financial context since options are best known in that setting. A call option is a financial tool which gives the option buyer the *right* to buy an asset at a certain price (called the strike price) at a certain time in the future (the exercise date). This right is not free, however; the buyer pays a premium to the seller in order to have the right to buy the underlying asset in the future. A put option is similar, but it gives the option buyer the right to *sell* an asset rather than buy.

I stress the words “right to buy” in the definition of option because that is the element which differentiates an option from a direct purchase. In no way is the option buyer obligated to buy the underlying asset; the buyer can choose not to buy the underlying asset, or in financial lingo, not to exercise the option. But why would an investor ever choose to not exercise? In general, if the strike price is more than the market price of the underlying asset at the time of exercising, the investor would lose money on the transaction. That is to say, the option buyer will only exercise the option if $MP > SP$, where MP is the market price and SP is the strike price. The premium is a sunk cost at this point, so any positive difference between the market price and strike price would make exercising worthwhile.

The most notable advantage of options over outright purchases is the reduction in risk. Since the buyer is not obligated to exercise the option, it shields the buyer from a drop in the value of the asset. A simple example will help illustrate this point. Imagine an investor has the choice to buy one share of Company X today at \$100 or an option to buy one share of Company X at \$100 in one year with an option premium of \$10. Suppose that the price of Company X's stock falls to \$50 over the course of the year. If the investor buys the share outright, then he has lost \$50. However, if the investor bought the option, he will choose not to exercise it since the market price (\$50) is less than the strike price (\$100), and he will only lose the premium of \$10. This example illustrates the ability of options to reduce risk.

Due to their ability to reduce risk, options are especially valued for volatile assets. The intuition is that if an asset is volatile, it has a greater probability of dramatically rising or falling in value. Since the option lessens the risk of a fall in value as illustrated above, the investor can reap the rewards of the volatile asset (i.e., the greatly rising price) while mitigating the risk of a drop in the asset price. In the case of an outright purchase, the investor must endure large losses which can result from volatility. Indeed, option premiums increase monotonically with increasing volatility as outlined in the Black-Scholes equation of option pricing, which implies that greater volatility makes options even more valuable (Day, n.d.). Conversely, options are an unnecessary purchase for less volatile assets (blue chips, e.g.) because their performance is easier to predict. Therefore, it is not worth paying the premium to reduce what little risk exists with less volatile assets.

2.2 MLB Contracts and the Labor Market

In order to explain how options play a role in baseball, I will first provide a brief explanation of MLB contracts and the labor market. Contracts are the means by which MLB teams (demanders of labor) hire players (suppliers of labor). A basic definition of a contract in this context is a legal agreement in which the team specifies for how many years and for what salary it will pay for the player's services. For example, a team may hire a player for only one year at \$X amount of dollars, or the team can hire a player for three years at \$X, \$Y, and \$Z dollars. The amount of money specified in each year is guaranteed. While length of contract and guaranteed salary are the basic elements of a contract, other elements can include performance and award incentives, signing bonuses, and even bonuses for losing weight. Despite the amount of other bells and whistles, the key contract figures are length of contract and guaranteed salary.

Not all contracts are treated the same, however. Starting in 1968, the Major League Baseball Players Association (MLBPA) has negotiated Collective Bargaining Agreements (CBAs) with teams in order to improve the players' contract outcomes. Although the newest CBA began in 2012, the focus of this paper will be the period covering the previous two CBAs, from 2003 to 2006 and 2007 to 2011. I pool the years falling under these CBAs because these years saw relative peace and prosperity from both the perspective of players and owners. In Major League Baseball, players fall into three different groups where each group is subject to different rules for determining salaries. The first group consists of players with less than three years of Major League service as determined by the rules of the CBA. These players can only negotiate with their first team, which essentially gives the team monopsony power. Naturally, players in this group are at a disadvantage when it comes to negotiations

because they cannot have their salaries bid up by competition from other teams. The second group consists of players who have between three and six years of Major League service. These players are known as “arbitration eligible” because they can take their demands for a higher salary to a neutral third party arbitrator. The arbitrator will then choose either the team’s figure or the player’s figure. However, many players negotiate a compromise salary with their teams before actually going to arbitration, which is a result of the player’s increased bargaining power. Arbitration eligible players still cannot negotiate with other teams, though. The third group of players is free agents, who have more than six years of Major League service. These players can sign with any team, giving them the greatest amount of bargaining power. This study will only focus on free agents, although the analysis could extend to arbitration eligible players as well.

The question is then: how are options used in MLB contracts? The most frequently used option is the team option, which gives the team the right to extend the player’s contract for another year at a predetermined price. Therefore, a team option appears to be very similar in definition to a call option: the team must choose to exercise the option by the last year of the guaranteed contract (exercise date), and if the option is exercised, the option specifies the price that the team must pay (strike price).

The analog of the option premium in MLB contracts is the buyout, which is a (usually small) amount of money paid to the player if the team chooses not to exercise the option. For example, if the team chooses not to exercise a \$1,000,000 option with a \$100,000 buyout on player X, the team will still have to pay \$100,000 to the player. The key difference between a buyout and a premium is that the buyout is only paid if

the option is *not* exercised, while a premium is paid regardless of the exercising decision. However, the fact that the buyout is a sum of money that the team must potentially pay the player makes it a premium.

Three other types of options exist in MLB contracts: player options, mutual options, and vesting options. Player options are essentially put options as described in the previous subchapter, as players buy the right to *sell* their services to a team. Mutual options require both player and team to exercise. Vesting options automatically exercise when a certain performance condition is met. For example, some options vest if the player plays in a certain number of games. Table 1 contains a list of frequencies of the number of contracts which contain each type of option. Club options were included in 17.09% of contracts for free agent hitters from 2003-2011, making it the most frequently bought type of option. The next most common type is mutual, which was only included in 2.70% of contracts. Therefore, we will ignore player and mutual options due to the small number of options awarded. Also, I will treat club and vesting options as equivalent. I contend that vesting options are essentially club options because the team would only offer a contract with a vesting condition such that it would be happy to rehire the player if the condition was met. Essentially, vesting options force teams to state their exercising conditions before the exercise date, while club options delay making decisions until the exercise date. However, both types represent the will of the team.

Table 1: Option Frequencies for Free Agent Hitters, 2003-2011

Type of Option	Frequency	Percentage of Contracts
No Option	423	76.08%
Club Option	95	17.09%
Player Option	11	1.98%
Mutual Option	15	2.70%
Vesting Option	12	2.16%

I have now drawn an analogy between financial options and MLB contract options, but it remains to be shown that teams use contract options efficiently. By efficient, I mean in the sense of financial options: that teams buy options from volatile players in order to reduce risk, and that they exercise options by comparing market value with strike price. I provide empirical evidence later in the paper to answer this question.

Chapter 3

LITERATURE

Despite the amount of research conducted on Major League Baseball salaries, few studies have asked questions about the use of options. Below I outline the literature which I reference in certain parts of my study.

In his Senior Thesis on the determination of pitcher salaries, Gross (2006) shows that years falling under different CBA periods need to be treated separately. This is because the properties of contracts can change at each new negotiation and therefore the determination of a player's salary can change as well. In that same study, Gross shows that pitchers and hitters need to be treated differently. This result is relatively self-evident as pitchers and hitters cannot be judged on the same statistics when determining value to the team. These findings will play a factor in selecting my sample.

Along the same lines of selecting the sample, Yosifov (2002) shows that free agent and arbitration eligible players need to be treated separately. As explained in the introduction, free agents are able to negotiate with any team while arbitration eligible players can only negotiate with their current team. As a result, free agents are in a better bargaining position and can command higher salaries, *ceteris paribus*. Therefore, the groups must be treated differently in the regression analysis.

In another Senior Thesis, Mooney (2008) provides an econometric model for salary determination which is relevant for my salary prediction model. In his model, he uses the natural log of real salary as the dependent variable. Real salary is the

nominal salary figure for a player adjusted for inflation. Adjusting for inflation is important in order to compare salary figures over a long period of time. The natural log of real salary is then taken because of the special interpretation of log values. When the dependent variable is in log form, then the coefficients of the independent variables can be interpreted as percentage changes in the dependent variable. This makes comparisons between different salary levels in different periods meaningful.

In the only paper I found which deals directly with the issue of MLB contract options, Dinerstein (2007) shows that teams have no discernible pattern in exercising club options. His data includes all hitters who played between 2001 and 2004. He runs a logistic regression where the dependent variable is whether an option is exercised or not. His only independent variable is called *PREDICTSAL – OPTIONCOST*, which is a predicted salary figure for the last year in a contract minus the cost of the option and buyout. He finds that the coefficient of *PREDICTSAL – OPTIONCOST* is statistically insignificant and thus has no effect on the probability of exercising an option, counter to financial option theory. I use Dinerstein's independent variable in a probit equation which estimates the factors impacting whether or not the option is exercised.

Chapter 4

DATA AND METHODOLOGY

4.1 Data

The majority of my data comes from Dr. Charles Link's and Dr. Daniel Brown's compiled dataset of baseball contract and performance details. The contract data (including team, date of signing, salary, bonuses, and length) comes from the MLB Joint Exhibit, which is not released to the public. The performance data comes from various public sources, including retrosheet.org. The data spans from 1984 to 2011, making it one of the most extensive baseball datasets ever assembled.

The dataset consists of 23,234 player-year observations. That is to say, a player who plays for five years appears as five different observations. However, the figure of 23,234 observations includes hitters and pitchers spanning over all levels of experience (no bargaining power, arbitration eligible, and free agent). As explained in the literature survey, hitters and pitchers must be treated separately, as well as players in different experience classifications. Therefore, I limit the sample for my study to free agent hitters who played in the 2003 season onward. I choose to examine free agent hitters because their market most closely resembles perfect competition and it is simpler to construct a measure of volatility for hitters as opposed to pitchers (to be explained in the next section). In terms of year constraints, 2003 to 2011 represents the last two CBA periods and can be grouped together due to labor peace during these years.

The final sampling constraints are different for the buying and exercising models (to be described in the next chapter). For the buying model, I limit target observations to those in the first year of a contract. By doing so, I analyze new contracts rather than player-years. The decision to buy an option occurs at the beginning of a contract, so we therefore focus on teams' decision making when signing new contracts. I also add the sampling condition for the option buying model that players must have greater than or equal to 130 plate appearances in the previous year. Plate appearances (PA) is a measure of how many times a player is either put out or reaches base, so it essentially measures how often a player is used. I choose 130 as a lower limit to PA in order to examine players who get a significant amount of playing time; 130 PA is the minimum to qualify for the Rookie of the Year award, so 130 is an objective minimum determined by MLB standards (Hakes and Sauer, 2006). By imposing these restrictions, the dataset for this model consists of 535 new contract observations. For the exercising model, we only want to examine players who we know have a club or vesting option in their contract. This condition limits the analysis to 108 contract observations. As explained in the next chapter, these models are estimated separately and therefore can have different sample sizes.

In terms of measuring offensive performance, I use on-base plus slugging (OPS). OPS is the sum of the on-base percentage and slugging average. On-base percentage is a measure of how often a player reaches base, taking into account hits, walks, and getting hit by a pitch. Slugging average is a measure of hitting power defined as the sum of total bases divided by at-bats. By adding on-base percentage and slugging average, OPS provides a holistic offensive metric which includes both hitting frequency and power.

Other key performance variables include PA, days spent on the disabled list (DL), and awards. PA measures how often a player is used, and hitters who play more are ostensibly more valuable to the team. The DL is a player designation which the team assigns when a player is injured, thereby allowing another player to temporarily take the injured player's roster spot. An oft-injured player might be perceived as riskier, so the number of days spent on the DL could play a role in determining whether a player has an option in his contract. Finally, player awards that reflect in-game performance include All-Star selections, Gold Gloves, and Silver Sluggers. Starting hitters are selected for the All-Star game based on fan votes. Although players are chosen by fans, the top performers tend to emerge regardless. Gold Gloves are awarded to players for fielding (defensive) prowess. Silver Sluggers are awarded to players in each league (American League and National League) for having the best offensive performance at a given position, where "best" is determined by a panel of coaches. Although all three of these awards are given based on somewhat subjective methods, they still signal outstanding player performances.

Other key control variables include length of contract, ethnicity, MLB experience, and current playing year. Length of contract is a particularly important control variable in the option buying model. It is difficult for teams to predict how a player will perform many years in the future, so the length of contract could be positively correlated with the probability of buying an option. That is to say, a longer contract leads to more uncertainty in the player's future performance and therefore could increase the probability of a team buying an option. I include various combinations of these variables in the models to reduce unexplained variance.

As described in the introduction, one should consider buying an option if an underlying asset is volatile. Therefore, I hypothesize that teams take player volatility into account when buying options. However, no standard measure of offensive volatility exists. Therefore, I constructed a variable named *STDV_OPS_LAG*, which is the standard deviation of monthly OPS lagged one year from the year of signing of the current contract. Since OPS is a measure of offensive efficiency, I found monthly OPS values for each player in a given year and took the standard deviation of those values. Standard deviation is a measure of spread, so a larger standard deviation means that a player is more volatile. I collected this data from the MLB.com player statistics database.

A list of shared performance and player characteristic variables used in the models is included below:

OPS_2YRAVG: 2-year average of OPS.

PA_2YRAVG: 2-year average of PA.

DAYS_ON_DL_3YRAVG: 3-year average of days spent on DL.

ALLSTAR_3YRAVG: 3-year average of All-Star appearances.

GOLDGLOVES_3YRAVG: 3-year average of Gold Glove awards.

SILVERSLUGGER_3YRAVG: 3-year average of Silver Slugger awards.

LOC: length of contract.

DLATIN2: dummy variable that equals 1 if the player is Hispanic.

DBLACK2: dummy variable that equals 1 if the player is Black

DiscreteExp: indicator variable that determines player's integer value of MLB experience in years.

YEAR: indicator variable that determines current playing year.

POSITION_ALT: indicator variable that determines player's defensive position.

Some of the variables listed above have different suffixes in the two models. In the option buying model, the suffix is “yos_lag” to indicate that the statistics are taken from the year prior to the year of signing of the current contract. In the option exercising model, the suffix is “cur” to indicate that the statistics are taken from the current playing year. I specify these time frames because option buying takes place at the beginning of a new contract, while option exercising takes place at the end of a contract.

4.2 Methodology

I use two, independent probits to test if teams use options efficiently. The first probit estimates the probability of a team buying a club or vesting option when signing a new contract (i.e., the probability a player receives an option in a new contract). The second probit estimates the probability of a team exercising the option given that a player has an option.

A probit regression has a binary dependent variable; in other words, the dependent variable can only take a value of 0 or 1. The basic probit model is shown below:

$$(1) \quad \text{Prob}(Y = 1 | X) = \Phi(X' \beta),$$

where Y is the outcome, X is a vector of independent variables, β is a vector of coefficient estimates, and Φ is the cumulative distribution function of the standard normal distribution. $X' \beta$ is a z-score, which gives a measure of where a statistic lies

relative to the mean (centered at 0). By computing Φ at a given z-score, we retrieve as an output a value between 0 and 1 representing the probability that $Y = 1$.

4.2.1 Option Buying Model

In the first probit, I estimate the probability that a team buys an option on a player's services when signing a new contract. I hypothesize that a team is more likely to buy an option if the player's performance is more volatile. To test this, I use the model stated below:

$$(2) \text{ Prob}(CLUB_OR_VEST) = \beta_0 + \beta_1(STDV_OPS_LAG) + \alpha P + \gamma C,$$

where $CLUB_OR_VEST$ is a dummy variable that equals 1 if a player has a club or vesting option in his contract, $STDV_OPS_LAG$ is the standard deviation of monthly OPS in the year prior to the year of signing of the current contract, P is a vector of performance variables, and C is a vector of player characteristic variables.

As described in the data section, $STDV_OPS_LAG$ is an estimate of a player's offensive volatility; it measures the spread in performance over the course of a year. A large value of $STDV_OPS_LAG$ implies a more volatile player based on statistics from the year prior to the year of signing of the current contract, which I hypothesize should increase the probability of the team buying an option for that player's services in the current year. Therefore, I hypothesize that the coefficient on $STDV_OPS_LAG$, β_1 , will be positive.

Summary statistics for the variables can be found in Table 2 below.

Table 2: Variable Summary Statistics for Option Buying Model

VARIABLES	(1) mean (sd)
club_or_vest	0.198 (0.399)
stdv_ops_lag	0.239 (0.153)
loc	1.613 (1.179)
ops_2yravg_yos_lag	0.757 (0.0888)
pa_2yravg_yos_lag	413.5 (156.4)
days_on_dl_3yravg_yos_lag	16.74 (19.58)
allstar_3yravg_yos_lag	0.0991 (0.200)
goldgloves_3yravg_yos_lag	0.0436 (0.156)
silverslugger_3yravg_yos_lag	0.0318 (0.119)
dlatin2	0.258 (0.438)
dblack2	0.196 (0.398)
discreteExp	9.120 (2.237)
Observations	535

4.2.2 Option Exercising Model

The second step of the independent probit model is to estimate the probability that a team exercises an option given that a player has a club or vesting option. In this model, I hypothesize that teams compare the value of the underlying asset and the option cost to determine whether to exercise the option. I use the model below to estimate this probability:

$$(3) \text{ Prob}(EXERCISED) = \beta_0 + \beta_1(\log RealSalaryDiff_buyout) + \alpha P + \gamma C,$$

where *EXERCISED* is a dummy variable that equals 1 if the player's option was exercised, *logRealSalaryOpDiff* is the difference between the natural log of real predicted salary and the natural log of real option cost minus real buyout cost, *P* is a vector of performance variables, and *C* is a vector of player characteristic variables.

I use the variable *logRealSalaryOpDiff* to represent the difference between the benefit and cost of exercising an option. I define the variable as follows:

$$(4) \logRealSalaryOpDiff = \ln(realPredictedSalary) - \ln(realOptionCost - realBuyoutCost)$$

The benefit of exercising the option is represented by *realPredictedSalary*. I define *realPredictedSalary* as the estimated salary that a player would earn if he were to sign a new contract in the current year, adjusted for inflation. This estimation is itself derived from an ordinary least squares (OLS) regression of various performance and control variables on real salary, evaluated at a player's current year performance values. Although we know a player's actual salary in a given year, that figure was determined in the past for a multi-year contract. Therefore, the actual salary figure might not accurately reflect the player's marginal value to the team, which is why I use a salary prediction which better reflects a player's fair market value. Results of the wage regression can be found in the appendix.

I define the cost of exercising the option as *realOptionCost - realBuyoutCost*. *realOptionCost* is the predetermined salary that the team must pay the player if the option is exercised (analogous to the strike price), adjusted for inflation. *realBuyoutCost* is the price of the buyout (analogous to the premium) which is only

paid if the team does *not* exercise the option, adjusted for inflation. I subtract the buyout cost from the option cost because in some cases, it will be cheaper for the team to exercise a player's option at a loss than to pay the buyout. As a numerical example, suppose a player's fair market value in the last year of his contract is \$9.5 million, while his option year salary would be \$10 million. By exercising the option, the team would lose \$0.5 million. However, suppose the player had a buyout of \$1 million. In this case, the team would lose \$1 million by *not* exercising the option. Since the team must either exercise or not exercise the option, it will exercise and take a smaller loss as a result. This analysis is presented in Table 3.

Table 3: Cost-Benefit Analysis of Exercising Option with Buyout

Decision	Estimated Benefit	Option Cost	Difference
Exercise	\$9.5 million	\$10 million	-\$0.5 million
Don't Exercise	\$0	\$1 million (buyout)	-\$1 million

By defining cost this way, I ensure that *realSalaryOpDiff* is positive if the team's most cost-effective move is to exercise the option. Finally, I take the natural log of the cost and benefit figures separately in order to better compare different salary levels. I hypothesize that the greater the value of *logRealSalaryOpDiff*, the greater the probability that the option is exercised. Summary statistics for the variables can be found in Table 4 below.

Table 4: Variable Summary Statistics for Option Exercising Model

VARIABLES	(1) mean (sd)
exercised	0.352 (0.480)
logRealSalaryOpDiff	-0.382 (0.610)
ops_2yravg_cur	0.773 (0.0990)
pa_2yravg_cur	435.9 (169.0)
allstar_3yravg_cur	0.105 (0.240)
goldgloves_3yravg_cur	0.0309 (0.133)
silverslugger_3yravg_cur	0.0617 (0.199)
days_on_dl_3yravg_cur	24.19 (30.26)
dblack2	0.185 (0.390)
dlatin2	0.287 (0.454)
discreteExp	8.861 (2.164)
Observations	108

Chapter 5

RESULTS

5.1 Option Buying Model

We now examine the results of the first step of the independent probit model: the option buying model. Regression results can be found in Table 5 below.

Table 5: Probability of Buying Club or Vesting Option on Player's Services

EQUATION	VARIABLES	(1) club_or_vest
club_or_vest	stdv_ops_lag	1.000** (0.443)
	loc	0.0879 (0.0706)
	ops_2yravg_yos_lag	0.517 (1.099)
	pa_2yravg_yos_lag	0.00191*** (0.000613)
	days_on_dl_3yravg_yos_lag	0.00457 (0.00374)
	allstar_3yravg_yos_lag	-1.101** (0.476)
	goldgloves_3yravg_yos_lag	-0.421 (0.502)
	silverslugger_3yravg_yos_lag	1.278* (0.668)
	dlatin2	0.0949 (0.172)
	dblack2	-0.0908 (0.197)
	7.discreteExp	0.317 (0.234)
	8.discreteExp	0.256 (0.247)
	9.discreteExp	0.113 (0.283)

10.discreteExp	0.0788 (0.279)
11.discreteExp	0.198 (0.276)
12.discreteExp	-0.0384 (0.232)
2003b.year	0 (0)
2004.year	-0.0570 (0.298)
2005.year	0.185 (0.295)
2006.year	-0.182 (0.316)
2007.year	0.479* (0.284)
2008.year	0.00457 (0.313)
2009.year	-0.547 (0.366)
2010.year	0.211 (0.293)
2011.year	-0.178 (0.309)
2.position_alt	0.629** (0.269)
3.position_alt	0.151 (0.304)
4.position_alt	-0.182 (0.300)
5.position_alt	-0.132 (0.304)
7.position_alt	-0.0328 (0.272)
Constant	-2.711*** (0.788)
Observations	535
Pseudo R-squared	0.112

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As explained in the data section, the key hypothesized explanatory variable is *STDV_OPS_LAG*, which is the standard deviation of a player's monthly OPS in the previous year. Since standard deviation is a measure of variance, then a large value of *STDV_OPS_LAG* means that a player's offensive performance varies greatly between

months. According to option theory, one should buy an option if the underlying asset is particularly volatile; therefore, I hypothesize that teams will buy an option on a player's services if the player's performance is volatile. The regression results in Table 5 show this effect.

The coefficient on *STDV_OPS_LAG* is 1.000, which means that a one unit change in *STDV_OPS_LAG*, ceteris paribus, increases the z-score of the probability of receiving a club or vesting option by 1.000. Moreover, the coefficient is significant at the 5% level and positive, which confirms the hypothesis that the more volatile a player's offensive performance, the more likely a team will buy an option for that player's services. However, according to the summary statistics for this model (Table 2), the mean of *STDV_OPS_LAG* for players in this sample is 0.2392744 with a standard deviation of 0.152612. Therefore, a more reasonable unit change in *STDV_OPS_LAG* would be 0.1 rather than 1. For a 0.1 unit change of *STDV_OPS_LAG*, the z-score of the probability of receiving a club or vesting option increases by 0.1000.

Since z-scores are just a relative measure of where a statistic falls on the standard normal distribution, the regression results are difficult to interpret in terms of actual changes in probabilities. Therefore, we must evaluate the z-scores at various levels of *STDV_OPS_LAG* holding all other variables constant. Table 6 shows the marginal effect of changing *STDV_OPS_LAG* on probability of receiving an option. The first sub-table shows the partial derivative of the probability of buying an option with respect to *STDV_OPS_LAG*, holding all other variables at their means. The second sub-table shows the predicted values of the probability of buying an option at various levels of *STDV_OPS_LAG*, holding all other variables at their means.

Table 6: Marginal Effects of *STDV_OPS_LAG* on the Probability of Buying a Club or Vesting Option

Partial Derivative of P(club or vesting option) with Respect to *STDV_OPS_LAG*

VARIABLES	(1) Marginal Effect
stdv_ops_lag	0.253** (0.112)

Values of P(club or vesting option) at Varying Levels of *STDV_OPS_LAG*

<i>STDV_OPS_LAG</i>	(1) Estimated Probability of Buying Option
0.05	0.154*** (0.0228)
0.10	0.165*** (0.0202)
0.15	0.176*** (0.0180)
0.20	0.188*** (0.0165)
0.25	0.200*** (0.0164)
0.30	0.213*** (0.0182)
0.35	0.226*** (0.0216)
0.40	0.240*** (0.0263)
0.45	0.254*** (0.0320)
0.50	0.268*** (0.0383)
Observations	535

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From Table 6, we see that a one unit change in *STDV_OPS_LAG* increases the probability of buying an option by 25.33%, or a 0.1 unit change in *STDV_OPS_LAG* increases the probability by 2.533%. Although the effect of volatility is positive and

significant on the probability of buying an option, it does not have a large impact on the probability of buying option. This can be seen in the second part of Table 6, which shows predicted values of the option buying probability at different levels of *STDV_OPS_LAG*, holding all other variables constant at their means. As *STDV_OPS_LAG* varies from 0.05 (very consistent performance) to 0.5 (very volatile performance), the estimated probability of buying an option increases from 15.4% to 26.8%. Although *STDV_OPS_LAG* clearly plays some role in the decision to buy an option, there are other factors influencing this decision.

Only a few of the other variables are significant in the buying model. *pa_2yravg_yos_lag*, the 2-year average of plate appearances starting from the year prior to the year of signing for the current contract, is positive and significant at the 1% level with a coefficient of 0.00191. The marginal effect of a one unit increase in *pa_2yravg_yos_lag* is 0.048% increase in the probability of buying an option, while the marginal effect of a 100 unit change (a more reasonable change in average PA than a one unit change) is a 4.8% increase in the probability of buying an option. This reveals that teams are more willing to buy an option for players who play a significant amount of time. This might occur because top players tend to play more and those same players are paid more. Therefore, teams might only justify the use of options for contracts which deal with large salary amounts.

The only other significant performance variable is *ALLSTAR_3YRAVG_YOS_LAG*, the 3-year average of All-Star appearances starting from the year prior to the year of signing for the current contract. The coefficient on this variable is -1.101 and is significant at the 5% level. This coefficient translates to a marginal effect of -2.79% on the probability of buying an option for a 0.1 unit increase

in *ALLSTAR_3YRAVG_YOS_LAG* (a 0.1 unit change is more realistic than a one unit change because the variable is an average taken over a three year period). Since hitters are chosen as All-Stars for their top (and usually consistent) offensive performance, then having a greater average value of All-Star appearances means that a player performs at a consistently high level. Therefore, All-Star appearances might signal to teams that a player is a less risky and not worthy of an option.

The only other significant variable is the position dummy for catcher, which is significant at the 5% level with a coefficient of 0.629. Since catcher is a dummy variable, then the interpretation of the coefficient is the effect of being a catcher on the probability of buying an option compared to the base group, which in this case is shortstop. The marginal effect of being a catcher on the probability of receiving an option is 19.27%, which means that a shortstop is 19.27% more likely to receive an option than a shortstop. I do not attempt to explain this effect.

The model has a pseudo- R^2 value of 0.1119, which means that only 11.19% of the variance in the probability of buying a club or vesting option is explained by the independent variables in the model. Therefore, much of the variance in buying options is unexplained by this model. We included observable performance and control variables, so I ascribe the remaining variance to unobservable player characteristics. For example, a team might take an option on a player that it feels is streaky, even if the player is statistically consistent. Indeed, baseball scouting in the past consisted primarily of qualitative descriptions of player ability, and this type of analysis still exists today in contrast to quantitative, sabermetric analysis. In addition, non-performance risks could induce a team to buy an option on a player's services. For example, a player with severe off-field issues (drunk driving, domestic abuse, etc.)

could lead a team to want an option on that player's contract: if his behavior gets out of hand, the team is not stuck with the player for another year. The problem with these variables is that they are not quantitative, and therefore not useful for econometric analysis. However, they could explain part of the variance in option buying decision.

5.2 Option Exercising Model

The second step of the independent probit model is to estimate the probability of a team exercising an option. Regression results for this model can be found below in Table 7.

Table 7: Probability of Exercising a Club or Vesting Option on Player's Services

EQUATION	VARIABLES	(1) exercised
exercised	logRealSalaryOpDiff	0.741** (0.351)
	ops_2yravg_cur	3.406 (2.683)
	pa_2yravg_cur	0.000272 (0.00128)
	allstar_3yravg_cur	-0.636 (1.322)
	goldgloves_3yravg_cur	-0.620 (1.180)
	silverslugger_3yravg_cur	0.932 (1.625)
	days_on_dl_3yravg_cur	-0.00735 (0.00655)
	dblack2	-0.0383 (0.484)
	dlatin2	0.114 (0.489)
	7.discreteExp	1.077* (0.572)
	8.discreteExp	0.135 (0.598)
	9.discreteExp	-0.00383

	(0.728)
10.discreteExp	-0.0618
	(0.664)
11.discreteExp	0.902
	(0.658)
12.discreteExp	0.631
	(0.550)
2003b.year	0
	(0)
2004.year	0.390
	(0.762)
2005.year	0.674
	(0.830)
2006.year	0.848
	(0.818)
2007.year	1.029
	(0.810)
2008.year	1.116
	(0.749)
2009.year	0.466
	(0.984)
2010.year	1.291*
	(0.737)
2011.year	1.759*
	(0.927)
2.position_alt	0.227
	(0.696)
3.position_alt	0.865
	(0.838)
4.position_alt	0.735
	(0.823)
5.position_alt	-0.0318
	(0.898)
7.position_alt	0.242
	(0.776)
Constant	-4.452**
	(2.101)
Observations	108
Pseudo R-squared	0.296

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The key hypothesized explanatory variable is *logRealSalaryOpDiff*, which is the difference between a player's predicted value (benefit to the team) and the option salary minus buyout cost (cost to the team). According to option theory, one should

exercise an option if the market value of the underlying asset is greater than the predetermined strike price. In this case, the market price is the player's predicted marginal revenue product to the team, which is estimated by a wage regression found in the Appendix. The strike price is the salary that the player would earn if the option is exercised minus cost of the buyout. For a discussion on why I include the cost of the buyout in the strike price, refer to Table 3 in the methodology section.

The coefficient on *logRealSalaryOpDiff* is 0.741, which means that a one unit increase in *logRealSalaryOpDiff*, ceteris paribus, increases the z-score of the probability of exercising an option by 0.741. Moreover, the coefficient on this variable is significant at the 5% level and positive, which confirms the hypothesis that the greater the difference between a player's fair market value and the cost of the option minus the buyout, the more likely a team will exercise the option. However, according to the summary statistics for this model (Table 4), the mean of *logRealSalaryOpDiff* for players in this sample is -0.382 with a standard deviation of 0.610. Therefore, a more reasonable unit change in *logRealSalaryOpDiff* would be 0.5 rather than 1. For a 0.5 unit increase of *logRealSalaryOpDiff*, the z-score of the probability of receiving a club or vesting option increases by 0.3705.

Since z-scores are just a relative measure of where a statistic falls on the standard normal distribution, the regression results are difficult to interpret in terms of actual changes in probabilities. Therefore, we must evaluate the z-scores at various levels of *logRealSalaryOpDiff* holding all other variables constant. Table 8 shows the marginal effect of changing *logRealSalaryOpDiff* on the probability of exercising an option. The first sub-table shows the partial derivative of the probability of exercising an option with respect to *logRealSalaryOpDiff*, holding all other variables at their

means. The second sub-table shows the predicted values of the probability of exercising an option at various levels of *logRealSalaryOpDiff*, holding all other variables at their means.

Table 8: Marginal Effects of *logRealSalaryOpDiff* on the Probability of Exercising a Club or Vesting Option

Partial Derivative of P(exercised) with Respect to *logRealSalaryOpDiff*

VARIABLES	(1)
	.
logRealSalaryOpDiff	0.258** (0.121)
Observations	108

Values of P(exercised) at Varying Levels of *logRealSalaryOpDiff*

VARIABLES	(1)
	.
-1.5	0.145* (0.0798)
-1.0	0.219*** (0.0664)
-0.5	0.312*** (0.0444)
0	0.419*** (0.0558)
0.5	0.533*** (0.101)
1.0	0.644*** (0.144)
Observations	108

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

From Table 8, we see that a one unit increase in *logRealSalaryOpDiff* increases the probability of exercising an option by 25.8%, or a 0.5 unit increase in *logRealSalaryOpDiff* increases the probability by 12.9%. The coefficient is positive and significant on the probability of exercising an option, which confirms the hypothesis that a larger positive difference between benefit and cost increases the probability of exercising an option. In addition, the coefficient of *logRealSalaryOpDiff* causes an impactful change in the dependent variable. This can be seen in the second part of Table 8, which shows predicted values of the option exercising probability at different levels of *logRealSalaryOpDiff*, holding all other variables constant at their means. As *logRealSalaryOpDiff* varies from -1.5 (cost far outweighs benefit) to 1 (benefit outweighs cost), the estimated probability of exercising an option increases from 14.5% to 64.4%. Therefore, teams appear to perform this cost-benefit analysis regularly when making exercising decisions and thus conform to option theory.

In fact, no other coefficient is significant at the 5% level in the model. This result indicates that *logRealSalaryOpDiff* encompasses many measures that teams might consider when determining a player's benefit. Controlling for offensive performance (OPS, PA, etc.) and other control variables (year, position, and ethnicity) does not make *logRealSalaryOpDiff* insignificant. This makes sense, as the benefit portion of *logRealSalaryOpDiff* is based on a wage regression which takes into account all current-year performance variables included in this model (see Appendix for more details). The only other coefficients significant at the 10% level are those for 7.discreteExp, which is an experience dummy for players with between 7 and 8 years of experience, and yearly dummies for 2010 and 2011. There does not appear to be any explanation for the significance of these coefficients. The positive coefficients on

2010 (1.291) and 2011 (1.759) could indicate a recent trend that teams are exercising options more frequently, but one would have to examine data for the years 2012 and onward to see if this result holds.

The model has a pseudo- R^2 value of 0.296, which means that 29.6% of the variance in the probability of exercising a club or vesting option is explained by the independent variables in the model. Therefore, some of the variance in exercising options is unexplained by this model. Similarly to the buying model, I ascribe the remaining variance to unobservable player characteristics. Two theories come to mind immediately: player popularity with the team and with the fans. If a player is popular with the current team, then the team will be more apt to exercise the option on that player's services. For example, the manager might like the effort the player shows on the field (effort being a qualitative, nebulous variable) or he might appreciate the player's presence in the locker room (veteran, high-moral player). Similarly, a player who is a fan favorite might have his option exercised more frequently than a little-known player. For example, if the player frequently participates in charity events or signs autographs at every game, fans will approve of the player more than an aloof player. More fans will pay to see the player which means more revenue for the team, and therefore the team might be more inclined to exercise that player's option. Both of these variables are qualitative and not easily measured, which poses difficulties for quantitative analysis. However, these reasons seem like plausible theories as to why some of the variance is unexplained.

Chapter 6

CONCLUSION

Teams adhere to financial theory to an extent when dealing with contract options. In terms of buying options, teams tend to be more unpredictable. I showed that teams take volatility into account when buying options, as a 0.1 unit increase in *STDV_OPS_LAG* increases the probability of buying a club or vesting option by 2.533%. However, the option buying model as a whole did not explain much of the variance in the actual option purchases (pseudo- $R^2 = 0.1119$). In terms of exercising options, teams are more predictable. I showed that teams compare the fair market value of the player to the cost (actual option salary minus buyout), as a 0.5 unit increase in *logRealSalaryOpDiff* increases the probability of exercising the option by 12.9%. This model explained far more of the variance in how teams actually exercised options (pseudo- $R^2 = 0.296$).

I propose that teams could use options more efficiently by making player volatility a more important factor when buying club or vesting options. Options are not free (premium for investors, buyout for baseball teams), so it does not make sense to buy options on assets which perform consistently well: the returns on such an investment are easily projected. Just as investors buy options on risky assets (high risk/high reward), teams should buy options on players who have the chance to provide high rewards but carry significant on-field risks as well.

However, teams might not be able to make volatility a more important factor because there is no standard metric of offensive volatility. To this end, an area of

future of research would be to find better measures of offensive volatility.

STDV_OPS_LAG only takes monthly OPS figures into account; a better measure would examine OPS over finer time intervals (weekly or biweekly, e.g.) or take multi-year averages of *STDV_OPS_LAG*. These statistics could provide teams with a more reliable means of determining which players are riskier and thus deserving of an option. In the long run, teams could save money by not buying options on players whose performance is predictable.

REFERENCES

- Day, T. (n.d.). "What Is the Implied Volatility for a Call Option." *Finance 6310*. Retrieved , from <http://www.utdallas.edu/~tday/6310FAQ3.html>
- Dinerstein, M. "Free Agency and Contract Options: How Major League Baseball Teams Value Players." Stanford University Senior Thesis (2007).
- Gross, A. "Salary determination of Pitchers in Major League Baseball." University of Delaware Senior Thesis (2006).
- Hakes, J.K. and R.D. Sauer. "An Economic Evaluation of the *Moneyball* Hypothesis." *Journal of Economic Perspectives* 20 (2006): 173-185.
- Mooney, D. "Moneyball and On-Base Percentage: Does Anyone Listen?" University of Delaware Senior Thesis (2008): 15-16.
- Yosifov, M. "Salary Determination in Major League Baseball-1998-2001." Master Thesis, Department of Economics at the University of Delaware (2002).

Appendix A

WAGE REGRESSION FOR EXERCISING MODEL

The option exercising model (chapter 4.2.2) requires a wage estimation model in order to determine the estimated player benefit for the variable *logRealSalaryOpDiff*. I run an OLS regression on the natural log of real salary using performance variables lagged one year prior to the year of signing. The regression results provide the coefficients which relate performance to salary. To calculate a player's estimated benefit to the team, I then multiply the coefficients obtained from the OLS regression with statistics from the current playing year to determine the estimated benefit in the current year. This calculation is necessary because the actual salary a player receives does not always represent his true market value. For example, the salary a player receives in year five of a contract is determined in year one, regardless of how the player performs in the year in between. Therefore, one must use current performance variables to obtain the fair market value. The regression results are below:

Table 9: Effect of Performance Variables on Log Real Salary

VARIABLES	(1) logRealSalary
ops_2yravg_yos_lag	3.241*** (0.287)
pa_2yravg_yos_lag	0.00338*** (0.000194)
days_on_dl_3yravg_yos_lag	0.00563***

	(0.00124)
allstar_3yravg_yos_lag	0.335**
	(0.156)
goldgloves_3yravg_yos_lag	0.251*
	(0.128)
silverslugger_3yravg_yos_lag	0.286
	(0.189)
dblack2	-0.0299
	(0.0672)
dlatin2	0.124**
	(0.0600)
7.discreteExp	-0.0769
	(0.0749)
8.discreteExp	-0.0471
	(0.0826)
9.discreteExp	-0.142
	(0.0899)
10.discreteExp	-0.170*
	(0.0912)
11.discreteExp	-0.314***
	(0.0957)
12.discreteExp	-0.331***
	(0.0741)
2003b.year	0
	(0)
2004.year	-0.117
	(0.0911)
2005.year	-0.0148
	(0.0926)
2006.year	-0.0531
	(0.0953)
2007.year	0.148
	(0.0947)
2008.year	0.117
	(0.0957)
2009.year	0.0852
	(0.0986)
2010.year	0.0780
	(0.0946)
2011.year	0.215**
	(0.0987)
2.position_alt	0.240**
	(0.0954)
3.position_alt	-0.240**
	(0.104)
4.position_alt	-0.0407
	(0.102)
5.position_alt	-0.107
	(0.103)
7.position_alt	-0.0159
	(0.0897)
2.ot	-0.0917

	(0.143)
3.ot	-0.00606
	(0.163)
4.ot	0.170
	(0.162)
5.ot	-0.225
	(0.187)
6.ot	0.257*
	(0.155)
7.ot	0.118
	(0.167)
8.ot	-0.181
	(0.172)
9.ot	-0.190
	(0.215)
10.ot	0.129
	(0.148)
11.ot	-0.354*
	(0.194)
12.ot	-0.0868
	(0.157)
13.ot	-0.234
	(0.173)
14.ot	-0.485***
	(0.185)
15.ot	-0.0835
	(0.169)
16.ot	-0.130
	(0.151)
17.ot	-0.249
	(0.160)
18.ot	0.0315
	(0.153)
19.ot	-0.0507
	(0.148)
20.ot	-0.347**
	(0.159)
21.ot	-0.149
	(0.144)
22.ot	-0.0959
	(0.164)
23.ot	-0.486**
	(0.239)
24.ot	-0.188
	(0.148)
25.ot	-0.257
	(0.165)
26.ot	-0.169
	(0.141)
27.ot	-0.197
	(0.175)
28.ot	-0.427**

	(0.209)
29.ot	-0.199
	(0.151)
30.ot	-0.444**
	(0.194)
Constant	3.958***
	(0.228)
Observations	916
R-squared	0.709

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1