12-2012

On the Mathematical Expectation of Beating the S&P 500

Darryl D. Walker
Indiana University of Pennsylvania

ON THE MATHEMATICAL EXPECTATION OF BEATING THE S&P 500

A Thesis

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Master of Science

Darryl D. Walker

Indiana University of Pennsylvania

December 2012

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Mathematics

We hereby approve the thesis of

Darryl D. Walker

Candidate for the degree of Master of Science

Yu-Ju Kuo, Ph.D
Associate Professor of Mathematics, Advisor

H. Edward Donley, Ph.D
Professor of Mathematics

John J. Lattanzio, Ph.D
Professor of Mathematics

Greg Wisloski, Ph.D
Associate Professor of Mathematics

ACCEPTED

Timothy P. Mack, Ph.D
Dean
School of Graduate Studies and Research

Title: On the Mathematical Expectation of Beating the S&P 500

Author: Darryl D. Walker

Thesis Chair: Dr. Yu-Ju Kuo

Thesis Committee Members: Dr. H. Edward Donley
Dr. Greg Wisloski
Dr. John J. Lattanzio

Two-thirds of professional money managers cannot earn a return on investment greater than the rate provided by the S&P 500 index. Portfolio selection in the face of an uncertain future requires the use of subjective probabilities about that future and the impact those events may have on the market. These subjective probabilities about unknown future events and unknown future impacts are difficult, if not impossible, to estimate. This problem can be circumvented. The profit from a trade is modeled as a Bernoulli random variable with a stop-loss parameter. Using the historical behavior of an exchange traded fund called the DIA and describing the relationship between the DIA and its underlying call options through regression, the mathematical expectation of the expression is empirically determined. By making only one key assumption about the market—that it will react to future events in a similar manner to its history-a system of trading call options is developed whereby the mathematical expectation of the return on investment is greater than the return provided by the S&P 500 index. The result is confirmed via simulation.

ACKNOWLEDGEMENTS

Much work and effort have gone into the completion of this paper. It represents the keystone in a multi-year project to complete the Masters Degree in Applied Mathematics in a part-time manner. I am grateful to so many people that have in some way given support. First to my dear sweet wife, Rebecca, whose support is unparalleled. Her steadfast encouragement and willingness to rearrange the household schedule to accommodate my academic pursuits was a necessary condition. Also to my daughters Mackenzie and Madison who sacrificed “daddy time” while I was working to become “the math man.” Their good natured humor is contagious.

To all of the professors who assisted in my course work I give the warmest gratitude. Dr. Kuo, Dr. Short, Dr. Zhang, Dr. Donley, and Dr. Lattanzio were instrumental. They also appreciated the complexities of working full time as a general manager for a large corporation, then going to graduate school after work and then somehow completing the workload before the next class. My compliments to their sense of discretion. I must also thank my committee members, Drs. Donley, Lattanzio, and Wisloski for their input and review.

Lastly, an additional note must be made for my committee chair Dr. Kuo. I cannot thank her enough for taking me on in this endeavor, on top of her already cumbersome duties. I appreciate her patience with me on this project along with her valuable insights. It took much longer than I expected and I am sure I have overtaxed her good nature. Thank you so much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PRELIMINARIES AND HISTORICAL PERSPECTIVE

Burton G. Malkiel states in his bestselling book: *A Random Walk Down Wall Street* that "…the market prices stocks so efficiently that a blindfolded chimpanzee throwing darts at the Wall Street Journal can select a portfolio that performs as well as those managed by the experts (Malkiel, 2007, p.17)." This statement refers to the fact that two-thirds of professional money managers cannot earn a return on investment greater than that of the unmanaged S&P 500 index (Malkiel, 2007). Succinctly put, an individual with no specialized training can open an account, invest money in a fund that mimics the S&P 500, forget about it, and beat two-thirds of the professionals. This astonishing fact presents the professional money manager with a success or failure option: The money manager either beats the S&P 500 or does not. Hence, the quest for consistent market beating returns is a popular one which hinges on investment selection and the outcome of that selection.

Viewing the outcome of the investment selection problem from a success or failure perspective leads to an interesting analysis of investing vis-à-vis the Bernoulli trial and the binomial distribution. The simplicity of the Bernoulli and binomial distributions combined with the notion of mathematical expectation have been of keen interest to the author since Dr. Short's mathematical statistics class in the fall of 2007. Together, they allow for a novel recasting of the investment selection problem. It is the intent of this thesis to introduce an analysis of investing based upon the Bernoulli trial and mathematical expectation. It develops a novel profit equation where an investment outcome is either a success or failure. The formulation of this equation is a new contribution to the field as it appears nowhere in the literature. The mathematical expectation of this profit equation is then determined using the Bernoulli distribution. Using the resulting equation, it will be demonstrated via Monte Carlo simulation how this equation can be used to trade call option contracts that earns a mathematically expected return that beats the long term rate of return on the S&P 500.

The objective of this thesis is twofold. The first objective is to model investing from an outcomes based perspective by considering an individual investment as a Bernoulli trial. That is, given an investment, that investment will terminate in either one of two mutually exclusive ways: profit or loss. By developing an investment payoff function that is appropriate for use with a Bernoulli trial, the mathematical expectation of the investment can be calculated. The individual investment can then be generalized into an n investment scenario by using the binomial distribution. A series of n investments will terminate with x profitable outcomes and $n-x$ losing outcomes. The second objective of this thesis is to outline the historical development of investment analysis in terms of expected future returns and risk. A few preliminaries are in order.

Background Information

The Bernoulli trial, the binomial distribution, and mathematical expectation are central to the analysis and their definitions are adapted from Hogg & Tanis (2006) and Marx (2001).

Definition 1: A Bernoulli trial is an event that has only one of two possible outcomes: Success or failure. The probability mass function of a Bernoulli trial is

$$f(x) = \theta^x(1-\theta)^{1-x}, \qquad x = 0 \text{ or } 1$$

where θ represents the probability of success. A series of n Bernoulli trials leads to the binomial distribution. The probability mass function of the binomial distribution is given by the expression:

$$g(x) = \frac{n!}{(n-x)!x!}\theta^x(1-\theta)^{n-x},\ x = 0,1,\dots,n.$$

Definition 2: Let $h(x)$ be the probability mass function of a discrete random variable x with sample space S. Then the summation $\sum_{x\in S} u(x)h(x)$ is the expected value of the function $u(x)$ and is denoted $E[u(x)]$.

The development of a specific $u(x)$ as a payoff function from an investment is fundamental and will be considered in more detail in Chapter 2.

Literature Review

There are two broad subjects that dominate the investing decision: expected future returns and the risks associated with those returns. Harry Markowitz offered the first systematic treatment of the risk-return tradeoff that each investor faces (Markowitz, 1952). Considering expected returns to be a desirable thing and variance of returns an undesirable thing, Markowitz developed his Expected Returns-Variance of Returns (E-V) portfolio selection model whereby investors maximize the expected return while minimizing the variance associated with that return. Expected returns are based upon subjective probabilities about future events and their impact on the investment. The manner by which these subjective probabilities are formed is complex and Markowitz admits not making an attempt to provide one (Markowitz, 1952). Markowitz defines risk is as the variance of expected returns.   At the same time, Roy (1952) developed a similar analysis but advised maximizing (relative to the variance) the positive difference between the expected return and some “disaster level” return that the investor considers important.

Since expected future returns are based upon subjective probabilities about future events and the impact those events may have on an investment, they are difficult to analyze rigorously. Both Hull (2006) and Wilmott (2001) have used the normal and log-normal distributions as a substitute for these subjective probabilities. However, the normal and lognormal distributions do not accurately reflect the frequency of large scale price movements that do occur (Taleb, 2007).

The appropriate measure of risk remains a much debated topic (Steinbach, 2001). Measures other than variance have been suggested including standard deviation, semi-variance, expected value of loss, and probability of loss (Rubenstein, 1999). However, a cursory glance at the current textbooks on portfolio management reveal that Markowitz’ mean-variance analysis is the standard (Elton, 2010; Reilly et al., 2008).

Historical Perspective

This paper investigates two new ideas in this field. First, the novelty is found in the development of a profit equation where the investment is modeled as a Bernoulli random variable with only two possible outcomes: success or failure. The profit equation presented appears nowhere in the literature and thus represents a new contribution to the field. This equation affords the advantage of sharply defining a successful investment and whether that standard of success was met. Moreover, recasting the profit of an investment as a success or failure option allows the usage of two well known probability distributions—the Bernoulli and binomial. The parameters of these distributions can be estimated using historical market behavior.

The second contribution follows directly from the first, namely that the mathematical expectation of the profit equation can be explicitly determined without the use of subjective probabilities concerning future events and their impact on a given investment selection. It is impossible to know what future events will be and how those unknown events will impact market behavior. Speculating about the likelihood of a multitude of unknown events introduces an arbitrary element in to the analysis of expectation. Eliminating that arbitrary speculation puts the analysis of expectation on firm theoretical ground.

Key Assumption

Regardless of an unknown future event, the market will move one way or the other. A brief look at history shows a range of events that, prior to their occurrence, would have been difficult to foresee. Within the last sixty years, there have been multiple wars, financial crises, technology booms, and real estate busts. The historical market reactions to these events are known and their magnitude and direction can be measured. Whatever the future event, it is assumed that the market will react to future events in a manner similar to its history.

CHAPTER 2

MODELING AN INVESTMENT AS A BERNOULLI VARIABLE

The Bernoulli Profit Equation

To model profit as a Bernoulli variable, a suitable mathematical expression must be developed. First, it is assumed that there are only two possible outcomes: profit or loss. Break-even outcomes are ignored because the likelihood of an investment outcome being exactly zero is nil. Let p denote profit in dollars, I denotes the size of the investment in dollars, r is the rate of return on the investment, k is a stop-loss parameter expressed as a percent of the original investment remaining after a decline, and x is a random variable. Then,

$$x = \begin{cases} 0, & \text{if there is a loss} \\ 1, & \text{if there is a profit.} \end{cases}$$

Since the objective is to capture the profit of an investment when there is a success, i.e, x=1, and also to capture what happens to the investment after a loss, x=0, the profit function can be determined as follows:

$$p(x) = I[(1+r)x + k(1-x) - 1] \tag{1}$$

Note the role that k performs in this equation. It is a parameter that represents the percentage of the original investment remaining in the event of a loss. It determines the point at which the investor would choose to terminate the investment if the investment begins to lose money.

Equation (1) represents a general payoff function in terms of success or failure and thus captures the Bernoulli property of an investment.

To demonstrate, suppose an investment is made and turns out to be a failure. Hence, we substitute x=0 into (1):

$$p(0) = I[k - 1].$$

The amount of the loss is mitigated by the stop-loss parameter k. Now suppose that the investment was a success, that is, $x = 1$. Then substituting $x = 1$ into (1) we have:

$$p(1) = Ir.$$

The amount of profit is simply the investment times the rate of return, as expected.

Determining the Mathematical Expectation of the Profit Equation

Recall the expected value that $E[u(x)] = \sum_{x \in S} u(x)f(x)$. Replacing $u(x)$ with $p(x)$ and substituting $f(x)$ with the Bernoulli distribution, we can derive $E[(p(x)]$.

Hence, the expected profit is measured by the following expression:

$$E[p(x)] = \sum_{x=0}^{x=1} I[\,(1+r)x + k(1-x) - 1]\theta^{x}(1-\theta)^{1-x}.$$

This simplifies to

$$E[p(x)\,] = I[(1+r)\theta + k(1-\theta) - 1]. \tag{2}$$

Equation represents a new contribution to this field and its analysis provides further insight into the notions of risk, return, and stop-loss.

Analysis of Equation (2) for θ, r, k. An analysis of equation (2) demonstrates the agreement between the popular notions of risk, return, and stop-loss and their representation by θ, r, and k respectively in (2). Since we are dealing with an expectation for profit from an investment, we will consider what relationships must hold between the parameters when $E[p(x)] > 0$. For simplicity throughout the analysis, we will assume that $I = 1$.

Since $E[p(x)] > 0$, i.e., $I[(1+r)\theta + k(1-\theta) - 1] > 0$, we solve for θ and get the following:

$$\theta > \frac{1-k}{1-k+r}. \tag{3}$$

Since k is expressed as a percentage of the original investment, $0 \leq k < 1$. Since r is the profit margin of a successful investment, $r > 0$. Thus, $1 - k + r > 0$.

If the investor is willing to lose the entire investment, i.e. $k = 0$, then (3) simplifies to $\theta > \frac{1}{1+r}$. That is, the probability of success must be greater than the inverse of one plus the rate of return on the investment. Since a lower θ implies a larger risk, (3) explicitly states the relationship between risk and return. A larger risk must be accompanied by a larger return if the positive expectation assumption is going to be met. This agrees with the intuitive notions of risk and return. Inequality (3) makes this relationship between risk (θ) and return (r) explicit within the context of (2).

An interesting special case is the idea of a 'sure thing' investment. Mathematically, a sure thing would be an investment with a 100% probability of success i.e., no risk. Under the assumption of positive expectations, (3) demonstrates that the sure thing is impossible.

Solving $E[p(x)] > 0$ for k yields

$$k > 1 - \frac{\theta}{1-\theta} r.$$

Since the future outcome of an investment is unknown, we can conclude that $\theta < 1$. The parameter k is a somewhat arbitrary in the following sense. It is the percentage of the original investment that we wish to retain if the value of the investment starts to decline. It represents a bail-out point if the investment starts to fail. In this sense, k can be interpreted as a measure of risk tolerance to the investor. Thus, the investor has some discretion in setting this value. In order to meet the positive expectation criteria, $k > 1 - \frac{\theta}{1-\theta} r$. The semi-arbitrary character of k is worth further investigation. The probability of success, θ, is linked to the future price movement of the investment. The direction of that movement is not known. Moreover, the magnitude of that movement will have an impact on r. Certainly the magnitude of a future price movement is not known. However, with a given probability of success and a given rate of return, k must be set within the range of $1 - \frac{\theta}{1-\theta} r < k \leq 1$ to guarantee a positive expectation on the investment. To illustrate this, we can set $E[p(x)] = 0$ and plot values for θ, k, and r where $E[p(x)] = 0$ is true. The result is graphically depicted by Figure 1.

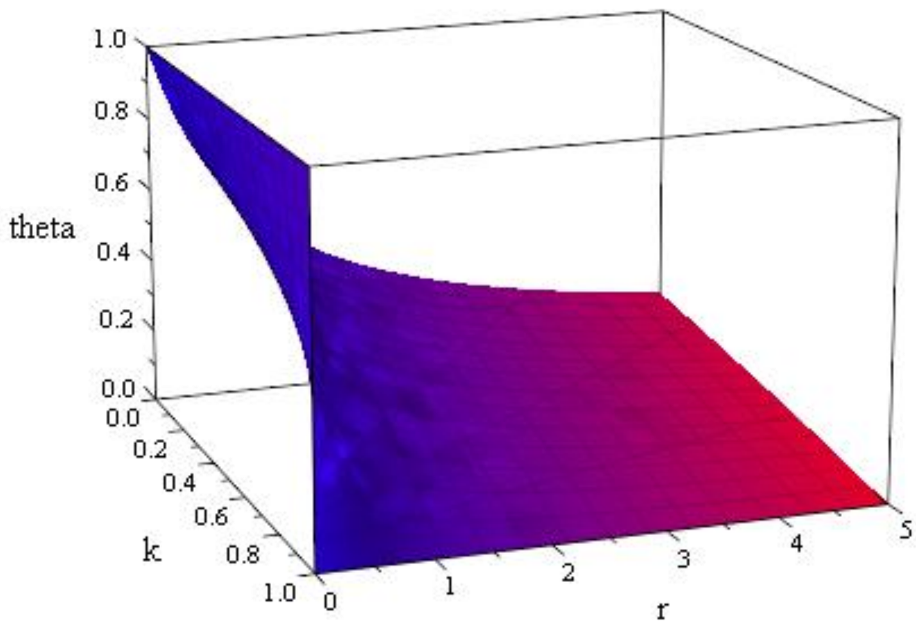

Figure 1. The surface between positive and negative expectations. This figure represents values of θ, k, and r where $E[p(x)] = 0$. Space above the surface has positive expectations; space below the surface has negative expectations.

It is now natural to consider two situations:

(i) Is there an investment vehicle where θ and r can be empirically estimated and then k determined to guarantee a positive expectation on the investment?

(ii) Is the expectation large enough to beat the long term annualized rate of return of the S&P 500?

CHAPTER 3

ESTIMATION OF θ AND *r*

This chapter will introduce the investment vehicle where we shall apply Equation (2)—equity call options on the Diamonds exchange traded fund. We will also determine what θ and r mean within the context of investing in these equity call options. Lastly, we shall analyze the data and estimate the value of θ and r. This will enable us to consider two things. First, is there a value of k that guarantees a positive expectation on the investment? And second, is that expectation greater than the long term rate of return on the S&P 500?

The Investment Vehicle: Equity Call Options

Equation (2) represents the expected profit from an investment. The investment can take many forms including stocks, bonds, options, antique cars, and real estate to name a few. In order to estimate the parameters θ and r, we must define the investment vehicle to be analyzed. Since the investment vehicle of this paper is equity call options on the Diamonds Trust exchange traded fund, the following definitions will be helpful.

Definition 3: The ETF. An exchange traded fund, known as an ETF, is a collection of individual stocks. The entire collection is assigned a symbol and traded just like an individual stock.

The Diamonds Trust ETF, symbolized by the letters DIA, is a collection of all of the stocks that comprise the Dow Jones Industrial Average. The Dow Jones Industrial Average is an index based upon 30 of the largest US-based publicly traded corporations. Thus the DIA is a collection of large, well-established companies and trades like an individual stock.

Definition 4: The Call Option. An equity call option is a contract that gives the option buyer the right, but not the obligation, to purchase a specified number of shares of stock at a specified price by a specified date. The specified price is called the strike price and the specified date is called the expiration date. The price of the option contract itself is called the premium.

Equity call options on the DIA will be the investment vehicle for this thesis. The elasticity of the call option is the reason for the choice of investment vehicle. Elasticity is the percentage change of the value of the option per 1% change in the price of the underlying asset which is typically between 8 and 10 (Reehl 2005). Since our objective is to have an expected profit greater than that of the S&P 500, the elasticity of options may serve as a better facilitator than individual stocks. Another kind of option, called a Put, may also be used. Put options confer upon the buyer the right to sell an underlying asset at a predetermined price and predetermined date. Where the price of call options rise with a rising price of the underlying asset, the value of a put option rises when the price of the underlying asset falls.

How simple call options work. Consider the January 110 DIA call option. This call option expires in January and has a strike price of $110 per share. Suppose the current price of the DIA is $112. Purchasing the $110 call confers to the buyer the right to purchase the DIA at $110 from the option seller. The option buyer can then sell these shares at the current market price of $112 thus realizing a $2/share profit. If the market price of DIA falls below $110 at expiration, then the option contract expires worthless. Hence, the value of the option contract on the DIA changes with the value of the DIA.

Establishing and Estimating r Using Regression

Recall that θ is the probability of a successful investment. Success in this context means that the option contract was profitable at the time of disposal. In other words, the rate of return, *r* in Equation (2) is positive and *r* refers to the rate of return on the equity call option. Since the value of the option contract depends on the magnitude and direction of the underlying asset price movement between the time of option purchase and the time of option disposal, the relationship between DIA price changes and option contract price changes must be determined.

Let r_c denote the rate of return on the option contract and C_t denote the premium of the option contract at time *t*. Then the equation

$$r_c = \frac{C_{t+1} - C_t}{C_t}$$

determines r_c.

Similarly, by letting r_a denote the rate of return for the underlying asset and a_t representing the price of the DIA at time *t*, the rate of return on the DIA at time *t* is given by:

$$r_a = \frac{a_{t+1} - a_t}{a_t}.$$

Since we wish to estimate *r* in Equation (2) then we must quantify the relationship between r_a and r_c. We will then use r_c as our estimate for Equation (2). Hence, we shall develop a scatter plot between r_c and r_a and then perform a linear regression.

There are three aspects of call option contracts that must be clarified before we proceed. The first is the strike price that will be analyzed. The second is the length of time that a contract will be held. Lastly, the time frame relative to expiration of the contract must be determined.

Strike Price Selection for the Regression. There are many different choices of strike prices for an options contract. Call options with strike prices that are below the current market price of the underlying asset are termed “in the money.” Likewise, options with strike prices at or near the current market price of the underlying asset are termed “at the money” while contracts with strike prices greater than the current market price of the underlying are called “out of the money” (Reehl 2005). For options on the DIA, the strike price changes by one dollar increments. Hence, one could choose to study the price movements for strike prices of $100, $101, $102,… or any value that is available for trade. For simplicity, we will look at three different strike prices representing options that are “in the money,” “at the money,” or “out of the money.” The first strike price will be the one closest to the closing price of the DIA. This will represent an “at the money” option. The second strike price will be two price levels above the closing price of the DIA. This will represent an “out of the money” option. The third strike price will be two price levels below the closing price of the DIA thus representing an “in the money” option. Thus, if DIA is trading at $110.32 on January 1, then the strike prices under consideration would be $108 (in the money option), $110 (at the money option), and $112 (out of the money option). Choosing three different strike prices will allow us to capture any strike price sensitivity in the relationship between r_a and r_c.

Holding Time of the Options Contract. Since the options contract will expire, a decision on how long to hold the option contract before disposal must be made. The length of time that a contract is held will allow the price of the underlying asset to rise or fall. This impacts the value of the option. For simplicity, we will use a holding period of five days which corresponds to one trading week in the market. This would allow for the taking of a position on Monday and disposing of it on Friday.

Time Frame Selection for the Regression. Since the option contract has an expiration date, its value is time sensitive (Hull 2006). There are many choices to be made in regards to the time frame studied relative to the expiration date of the option. Since we have chosen to hold an options contract for five days, we will study the four five-day time frames prior to expiration. That is, we will analyze the relationship between $\boldsymbol{r_c}$ and $\boldsymbol{r_a}$ with the time frames of 0-5 days prior to expiration, 6-10, 11-15, and 16-20 days prior to expiration. This will allow us to see time sensitivity in the relationship between $\boldsymbol{r_c}$ and $\boldsymbol{r_a}$.

There were two sources of data for the regression. The closing prices for the DIA were downloaded from the Yahoo! Finance website for the year 2009. A CD with option pricing was purchased from DeltaNeutral.com and spans the same time period. Regressions were performed using Microsoft Office Excel 2007.

Regression Results: The Proxy for r.

Analyzing the four different expiration time frames and three different strike prices yields twelve total regressions on the data. An inspection of the scatter plots suggests the linear model to be an appropriate fit to the data (See Appendix A). Thus, the regression equation is of the form $r_c = \alpha + \beta r_a$. Since we are interested in the coefficient of r_a, we perform the regression forcing $\alpha = 0$ and so the final model is of the form $r_c = \beta r_a$.

The objective in performing the regression is to capture the change in the option call price relative to the change in the asset DIA. The coefficients of r_a in the regression capture this information and are presented in Table 1 which follows. Complete regression results can be found in the Appendix A and Appendix B.

Table 1

Regression Coefficients

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	46.10	31.05	24.36
6 to 10	30.27	24.44	19.32
11 to 15	21.04	16.61	13.84
16 to 20	18.40	15.03	12.39

The coefficients listed in the table represent the percentage change in the bid price of the call option premium per 1% change in the price of the DIA. For example, a call option contract that is 6 to 10 days from expiration and two price levels out of the money can be expected to increase by 30.27% for every 1% change in the price of the DIA over the ensuing five days.

Furthermore, Table 1 illustrates two very clear trends. First, the out-of-the-money strike prices are more responsive to changes in the DIA than are the at-the-money strike prices; and at-he-money strike prices are more responsive than in-the-money strike prices. Purchasing out-of-the-money calls offers a greater opportunity of price appreciation for a given change in the DIA over at-the-money or in-the-money options.

The second trend appears relative to the time to expiration. The closer the option is to expiry, the greater the responsiveness of the call price to a given change in the DIA.

In choosing the value of r to use in Equation (2), one must consider how much time is left until expiration on a particular contract and whether the contract is out-of-the-money, at-the-money, or in-the-money at the time of purchase. Given these two choices, the appropriate value for r can be determined from Table 1.

Estimating θ

Recall that θ is the probability of success in a Bernoulli trial. Success in investing occurs when an investment is profitable. If one purchases an equity call option, it becomes profitable when the underlying asset moves up in price. Hence, a success entails that the DIA moves up during the investment holding period. Since the magnitude of the move influences r, success can be defined on several orders of magnitude. This section will analyze percentage moves in the DIA at 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, and greater than 3.0% that occur over a five-day period. For example, if the DIA closes up 0.5% sometime over the following five days then the trial is considered a success at the 0.5% level, but a failure at the other levels. The proportion of successes at a given magnitude as a percentage of total trials is the estimate for θ at that magnitude. Table 2 presents the results from 2009 on the percentage moves of the DIA and their associated success rates.

Table 2

Market Moves of a Successful Magnitude

Magnitude	>0.5%	>1.0%	>1.5%	>2.0%	>2.5%	>3.0%
Successes	168	135	117	92	77	57
θ	0.677	0.544	0.472	0.371	0.31	0.23

N=218

Thus, on 168 occasions, the DIA closed up at least 0.5% sometime over the following five days. It follows that 168/248=0.677 is the estimate for θ at the 0.5% magnitude.

The Long Term Rate of Return on the S&P 500

The ultimate purpose of this study is to determine if Equation (2) can assist an investor in securing a rate of return that is greater than the long-term annual rate of return on the S&P 500 index. This index was 17.56 on January 3, 1950 and closed at 1115.60 on December 31, 2009. Using the known continuously compounding equation $FV = PVe^{rt}$ where FV is the future value, PV is the present value, r is the annual interest rate, and t is the time in years, we can rearrange terms to solve for r.

Using 1115.60 as the FV and 17.56 as PV and letting t=60 (the number of years over which the return is calculated), we see that the annualized rate of return on the S&P 500 for the 60 years between January 3, 1950 and December 31, 2009 is 6.92%. This will serve as the benchmark against which we shall measure the results of Equation (2). If Equation (2) results in an expectation that is greater than 6.92% then it is successful; and a failure otherwise.

It has been noted that success in general can be determined on different levels of magnitude. To determine if Equation (2) is helpful in beating the S&P 500, we must determine the percentage change in the DIA that induces a change in the option contract of at least 6.92%. Since the regression results suggests that $r_c = \beta r_a$ it follows that $r_a = \frac{r_c}{\beta}$. Letting r_c=6.92 and dividing this by the regression coefficients from Table 1 we can determine the necessary changes in the DIA that induce the requisite magnitude change in the options contract that will exceed the long term rate of return of the S&P 500. These results are listed in Table 3 for the twelve different contracts studied.

Table 3

Minimum Percentage Move

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	0.150	0.223	0.284
6 to 10	0.229	0.283	0.358
11 to 15	0.329	0.417	0.500
16 to 20	0.376	0.460	0.559

Given the coefficients from Table 1, very little movement is necessary in the underlying asset DIA to induce a movement on the call option that will exceed the rate of return on the S&P 500.

With each magnitude listed in Table 3, there is a corresponding θ that indicates the success rate at which the DIA moves beyond the listed magnitude. These success rates are listed in Table 4. They are calculated in the same manner as Table 2. To illustrate this, consider an options contract that is out-of-the-money and 0-5 days from expiration. From Table 3 we know that the DIA must appreciate at least 0.15% in order for the option contract to appreciate at least 6.92%. Table 4 shows that that the DIA will appreciate at least 0.15% at a success rate of 0.774.

Table 4

Success Rates Associated with Minimum Moves

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	0.774	0.762	0.75
6 to 10	0.762	0.75	0.73
11 to 15	0.742	0.726	0.677
16 to 20	0.73	0.694	0.665

Having now determined the relationship between DIA price movements and the associated options contracts as well as the success rates associated with the magnitude of DIA price movements, we can now substitute these values into Equation (2) to get some results. In particular, we shall calculate the expected profit per trade for each contract under study. Further, since Equation (2) represents a single investment scenario, we will generalize that result into an n-investment scenario to determine the theoretical possibility of beating the S&P 500.

CHAPTER 4

RESULTS AND GENERALIZATION.

Chapter 4 will use the estimates for θ and r developed in chapter 3 to calculate $E[p(x)]$ at various levels of k. Since multiple trades can be made, it will also generalize $E[p(x)]$ from a single investment case to an n-investment scenario.

Calculation of $E[p(x)]$.

Recall that Equation (2) is given as:

$$E[p(x)] = I[(1+r)\theta + k(1-\theta) - 1],$$

and that under a positive expectation the parameter k must satisfy the following condition:

$$k > 1 - \frac{\theta}{1-\theta} r.$$

Using the values of θ from Table 4 and substituting 6.92% for r, the minimum value for the parameter k that guarantees a positive expectation can be determined for each option contract under study. This is presented as Table 5.

Table 5

Minimum Values of k

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	0.763	0.778	0.792
6 to 10	0.778	0.792	0.813
11 to 15	0.801	0.817	0.855
16 to 20	0.813	0.843	0.863

If we set k = 0.9 we can guarantee a positive expectation on all contracts under study, and calculate Equation (2) for each contract. These calculations are exhibited as Table 6:

Table 6

Expected Value of a Single Trade

Days to Expiration	Out Money	At Money	In Money
0 to 5	0.031	0.029	0.027
6 to 10	0.029	0.027	0.024
11 to 15	0.026	0.023	0.015
16 to 20	0.024	0.017	0.013

Thus, for each contract under consideration, we can set $k = 0.9$ and guarantee a positive expectation on an individual investment. Note that each contract's expectation is well below the long-term rate of return on the S&P 500 which is 6.92%. In order to exceed this value, multiple investments must be made. This requires a generalization of Equation (2) to handle multiple trades.

The n-Investment Scenario

We can set up a series of investments using the profit equation (1) and taking advantage of the Bernoulli property to generalize into an *n*-investment scenario. Note the following scheme:

$$P_1 = (1+r)x_1 + k(1-x_1) - 1$$

$$P_2 = (1+r)x_2 + k(1-x_2) - 1$$

$$\vdots$$

$$P_i = (1+r)x_i + k(1-x_i) - 1$$

$$\vdots$$

$$P_n = (1+r)x_n + k(1-x_n) - 1.$$

In this scheme, P_i is the profit dollars per trade and x_i is a Bernoulli random variable; *r* is the target profit rate. The total profit on *n* investments is the sum of all the individual trades. Denoting the total profit by T, we obtain

$$\begin{aligned} T &= \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} [\,(1+r)x_i + k(1-x_i) - 1] \\ &= \sum_{i=1}^{n} [x_i + rx_i + k - kx_i - 1] \\ &= (1+r)\sum_{i=1}^{n} x_i + k(1 - \sum_{i=1}^{n} x_i) - n \end{aligned}$$

If we set $y = \sum_{i=1}^{i=n} x_i$, then

$$T = y(1+r) + k(n-y) - n.$$

By definition, y represents the number of successes in *n* investing trials. Thus, *T* is a function of binomial random variable *y*.

Since y is a binomial random variable, we use the binomial distribution in the calculation of E[T(y)]. This produces the following expression:

$$E[T(y)] = \sum_{y=0}^{n} \frac{n!}{(n-y)!y!} \theta^{y}(1-\theta)^{n-y}[y(1+r) + k(n-y) - n]. \qquad (4)$$

Note that $\sum_{y=0}^{n} \frac{n!}{(n-y)!y!} \theta^{y}(1-\theta)^{n-y} y = E(y) = n\theta$;

$$\sum_{y=0}^{n} \frac{n!}{(n-y)!\,y!} \theta^{y}(1-\theta)^{n-y} y\,(n-y) = E(n-y) = (1-\theta)n;$$

and $\sum_{y=0}^{n} \frac{n!}{(n-y)!y!} \theta^{y}(1-\theta)^{n-y} = 1$.

Thus, Equation (4) reduces to

$$E[T(y)] = n[(1+r)\theta + k(1-\theta) - 1] \qquad (5).$$

Equation (5) tells us that the total expected profit for n trades per year is simply n times the expected profit per trade.

Dividing 6.92% by each value in Table 6 would generate the minimum number of investments per year necessary to equal the return on the S&P 500. Since there cannot be a partial investment in the context of this paper, each value is rounded up to the nearest whole number. An additional trade is added in order for the expectation to exceed the value of the S&P 500. These results are presented in Table 7.

Table 7

Minimum Number of Trades

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	4	4	4
6 to 10	4	4	4
11 to 15	4	4	6
16 to 20	4	5	7

The number of trades per year times the expected profit per trade yields the expected total profit for the year. Multiplying the values in Table 7 with the individual expected profit per trade in Table 6 yields Table 8.

Table 8

Expected Value of Multiple Trades per Table 7

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	12.38	11.57	10.76
6 to 10	11.57	10.76	9.41
11 to 15	10.22	9.14	8.32
16 to 20	9.41	8.71	8.76

Hence, it would appear that it is theoretically possible to use an investment strategy of trading call options and expect to beat the long-term rate of return on the S&P 500.

CHAPTER 5

SIMULATING THE STOCK MARKET

The results presented in chapter 4 rely on historical stock market data. The parameters of Equation (2) and the results of Equation (5) are based entirely on the given historical data. As such, it appears that trading call options can result in a mathematically expected rate of return that exceeds the S&P 500. But what about the future returns? Can the results of chapter 4 hold against new data coming in from an unknown future? It is for this purpose that a simulation of stock market movements is presented. The outcome of the simulated market behavior tests the results of chapter 4. Two scenarios need to be examined. The first is determining the result of a buy and hold strategy simulating the S&P 500 index. The second is determining the results of applying Equation (5) to the twelve different contracts under consideration. The two results will be compared to determine if the expectation of trading in call option contracts is greater than the return on the S&P 500. The bootstrapping method will be presented and its application to stock market data.

Monte Carlo Simulation and Bootstrapping.

Monte Carlo simulation involves drawing samples of data from a pseudo population and then studying the statistical properties of the sample. If the pseudo population has the pertinent characteristics of the real population under consideration, then the statistical properties of the pseudo population also represent the real population. The most important aspect of simulation is ensuring that the pseudo population has the properties of the real population (Mooney, 1997). The bootstrapping technique guarantees that this condition is met.

Bootstrapping is the process of drawing random elements from a data set. Each data point is assigned an index number starting at 1 with the first data point and ending with *n*, where *n* is the total number of data points in the set. A random number between 1 and *n* is generated. The data point mapped to this random number is selected as the simulated outcome. Bootstrapping thus captures the frequency characteristics of the data set (Winston, 2004).

Simulating the stock market using these methods allows us to generate multiple trials by which to test Equation (5). The historical data used to establish the parameters of (5) represents a single observed trial covering sixty years. But suppose we were to go back to 1950 and rerun history. Would the S&P 500 still generate the long term rate of return of 6.92%? Is the particular result that we got from history typical? Since we are moving forward in time into an unknown future, testing Equation (5) via simulation could determine its robustness.

All of the following simulations were performed using Microsoft Office Excel 2007.

Simulating the Annual Long Term-Rate of Return for the S&P 500.

In calculating the long-term annual rate of return on the S&P 500 of 6.92%, the time period encompassed January 3, 1950 to December 31, 2009. The data set for the simulated return uses the market behavior covering this time period. The following procedure outlines the process:

1. Gather the weekly opening and closing prices of the S&P 500 index from January 3, 1950 to December 31, 2009.
2. Calculate the weekly percentage change in the index using the closing price for the week and the opening price for the same week.

3. Assign a number to each weekly return, assigning the number 1 to the week of January 3, 1950, assigning the number 2 to the week of January 10, 1950, and so on, until the number 3130 is assigned to the week of December 31, 2009.
4. Use the Excel *randbetween* function to generate a random number between 1 and 3130. Find the weekly percentage change mapped to that random number. This represents a simulated market outcome.
5. Start a counter at the number 1 and update the counter by multiplying it by 1 plus the result from Step 4. For example, if Step 4 produces a market gain of 3.0%, then the value of the counter after the first week is 1.03. If on the next week, Step 4 produces a market loss of 2.0%, then the value of the counter at the end of week 2 is 1.03 x (0.98) = 1.009 (result rounded to three decimal places). The counter represents the simulated cumulative value of the S&P 500 index.
6. Repeat Step 4 and Step 5 a total of 3120 times, which is the number of weeks in a 60-year time period. This is the same length time period used in calculating the historical long-term rate of return in Section 3.4. Note the value of the counter. This is the simulated future value of the S&P 500 index after a 60-year time period. Compute the annualized rate of return from the value of the counter.

Running this simulation 1000 times produces an average annual return of 8.44%, with $\sigma = 1.96$. This represents 1000 trials of a 60-year time period given the historical weekly behavior of the S&P 500. The 8.44% will serve as the benchmark against which to measure the results of trading in options.

Simulating the Call Option Strategy

Simulating the call option strategy is similar to Section 5.2 but with some key differences. The historical basis for trading the DIA covers 573 weeks from January 1, 1999 to December 31, 2009. The weekly movements of the DIA during this time are assigned an index number between 1 and 573 and a random number between 1 and 573 is generated to determine a simulated outcome in the same manner as Section 5.2. This outcome is used as the input for the regression equations for each option contract. The coefficients of the regression equations that were listed in Table 1 of Section 3.2.4 are then used to determine the profit or loss from each option contract.

Since the call option strategy involves individual trades, the profit and loss for each trade is summed for a grand total net profit or loss. To track the overall return on our strategy, we start with $1. This represents a single dollar that we begin trading with. Add or subtract from that dollar the profit or loss from trading. At the end of a trading period, the sum represents the total profit or loss including the original starting dollar. The rate of return on that series of trades is the ending value of the sum minus the original dollar.

Recall that Table 7 listed the minimum number of trades for each option contract necessary to exceed the return on the S&P 500. The call option simulation used that many trades for each option contract in order to have a relevant comparison with the results of chapter 4. In other words, from Table 7, the out-of-the money call option contract with 0 to 5 days left to expiration required a minimum of four trades to exceed the return on the S&P 500. The simulated result of the call option strategy used four trades for that particular contract. The other contracts used the same number of simulated trades as listed in Table 7.

Here is a step by step procedure for this simulation:

1. Collect the opening and closing prices of the DIA from January 1, 1999 to December 31, 2009.
2. Calculate the weekly percentage change in the DIA using the closing price for the week and the opening price for the same week.
3. Assign the number 1 to the first week, assign the number 2 to the second week, and so on, until the number 573 is assigned to the week ending December 31, 2009.
4. Use the Excel command *randbetween* to generate a random number between 1 and 573. Find the weekly percentage change in the DIA mapped to the random number.
5. Start twelve counters at $1, each counter representing one of the twelve different option contracts under study.
    a. If the result from Step 4 is positive, multiply the coefficients in Table 1 by the result from Step 4 and add the product to the corresponding counter for each contract.
    b. If the result from Step 4 is negative, subtract 1-k from the counter for each contract. For simplicity, the results that follow assume that k = 0.9. In the event that the loss is less than 10%, the final result will be greater profit. Accounting for losses with k=.9 provides a more conservative estimate.
6. Repeat Step 4 and Step 5 for each contract according to the frequency listed in Table 7.
7. Subtract 1 from the final value of each counter. This total represents the rate of return on the option contract.

The simulation process outlined above was repeated 1000 times and Table 9 below presents the average result:

Table 9

Simulated Results of Option Trading

Days to Expiration	Out of the Money	At the Money	In the Money
0 to 5	1.590	0.961	0.694
6 to 10	0.980	0.753	0.554
11 to 15	0.614	0.442	0.511
16 to 20	0.515	0.480	0.492

To illustrate, consider the out-of-the-money calls that are 16 to 20 days from expiration. From Table 7 of Section 4.2, this trade must be made four times per year in order to have an expected return greater than the long-term rate of return on the S&P 500. The simulated result listed in Table 9 for this particular contract is 51.5%. Considering all of the contracts under study, the minimum average return is 44.2% while the maximum is 159%. The simulated return to the S&P 500 was only 8.44%. It would appear that trading in call options is further supported by these simulated results.

Discussion and Conclusion

There are three main areas that demand further discussion. First, the simulated returns presented in Table 9 do not take into consideration the cost of brokerage fees. These would have to be included in order to get a more realistic result. There are trading costs per options contract as well as a fee to initiate the trade. Typically, the fees range from $4.95 to $9.99 to initiate a trade and an additional $0.65 to $0.99 per contract commission. Thus, the returns listed in Table 9 are higher than they would be with the fees included. The impact the brokerage fees have on these returns can be mitigated by volume, as the fixed cost of initiating a trade is spread over a large dollar amount. Nevertheless, even without the brokerage costs, the minimum average return of 44.2% from Table 9 offers a lot of excess profit over the 8.44% in order to pay for those fees.

Secondly, the results depend on the regression coefficients presented in Table 1. The data for the regressions represents one year—2009. The behavior of option contracts relative to market fluctuations may not be representative during that year. More years included in the regression would determine if the results used in these regressions are typical. Unfortunately, option data is expensive. The data CD for 2009 was purchased at a cost of $250. Moreover, extracting the relevant data from thousands of daily quotations is a very time consuming process. However, this would be a fertile area for future research.

Lastly, prices move quickly in the market. The simulated results in Table 9 as well as the theoretical results from Table 8 assume a stop-loss parameter of 0.9. Underlying the results of both tables is the idea that an investor would be nimble enough to close out a position before the market generates more losses than intended. It is unlikely, given the volatile relationship between the market move and the move in the corresponding option that the investor would be nimble enough to actually execute at the 0.9 value. This, however, may be overcome by placing more trades per year than the maximum used in the simulation.

The simulated results of chapter 5 use a very small number of trades. The maximum number is seven. Since the regressions are based upon a five-day time period, it is possible to place 52 trades per year. This would provide a mathematically expected return far greater than the returns calculated here.

Lastly, it should be mentioned that while the future holds infinitely many possibilities, history unfolds uniquely. The techniques used here depend on a large number of trials and the averages taken as a proxy for what is most likely to happen. Once a particular trade is placed the particular result could be in violation of the averages. In the simulated results, there were several instances where all trades for the simulated year were losers. In reality, this run of "bad luck" could cause an investor to quit the market altogether thus losing any chance at beating it.

The original problem is the observation that two-thirds of professional money managers cannot beat the long-term rate of return on the S&P 500. By modeling the profit from a trade as a Bernoulli random variable and then using market behavior to estimate the parameters of the equation, we have determined that it is theoretically possible to trade options and earn a mathematically expected return that is in excess of the S&P 500. Subjective probabilities concerning the impact of future unknown events on the market, problematic from Markowitz' famous 1952 paper to the current day, are altogether abandoned.

References

Hogg, R. V., & Tanis, E. A. (2006). *Probability and statistical inference 7th ed*. Upper Saddle River, NJ: Prentice- Hall.

Hull, J.C. (2006). *Options, futures, and other derivatives 6th ed*. Upper Saddle River, NJ: Prentice-Hall.

Malkiel, B. G. (2007). *A random walk down Wall Street*. New York, NY: W.W. Norton.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7* (1), 77-91.

Marx, M. L.,& Larsen, R.J. (2001). *An introduction to mathematical statistics and its applications 3rd ed*. Upper Saddle River, NJ: Prentice-Hall.

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Reehl, C.B. (2005). *The mathematics of options trading*. New York, NY: McGraw-Hill.

Reilly, F. K., & Brown, K. C. (2000). *Investment analysis and portfolio management 6th ed*. Orlando, FL: Harcourt.

Roy, A.D. (1952). Safety first and the holding of assets. *Econometrica*, *20*, 431-449.

Rubenstein, M. (2002). Markowitz's portfolio selection: A fifty year retrospective. *The Journal of Finance*, *62* (3), 1041-1045.

Steinbach, M. C. (2001). Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review*, *43* (1), 31-85.

Taleb, N. N. (2007). *The black swan*. New York, NY: Random House.

Wilmott, P. (2001). *Paul Wilmott introduces quantitative finance*. West Sussex, England: John Wiley.

Winston, W. L. (2004). *Microsoft Excel: Data analysis and business modeling*. Redmond, WA: Microsoft Press.

Appendix A Plots for Selected Regressions

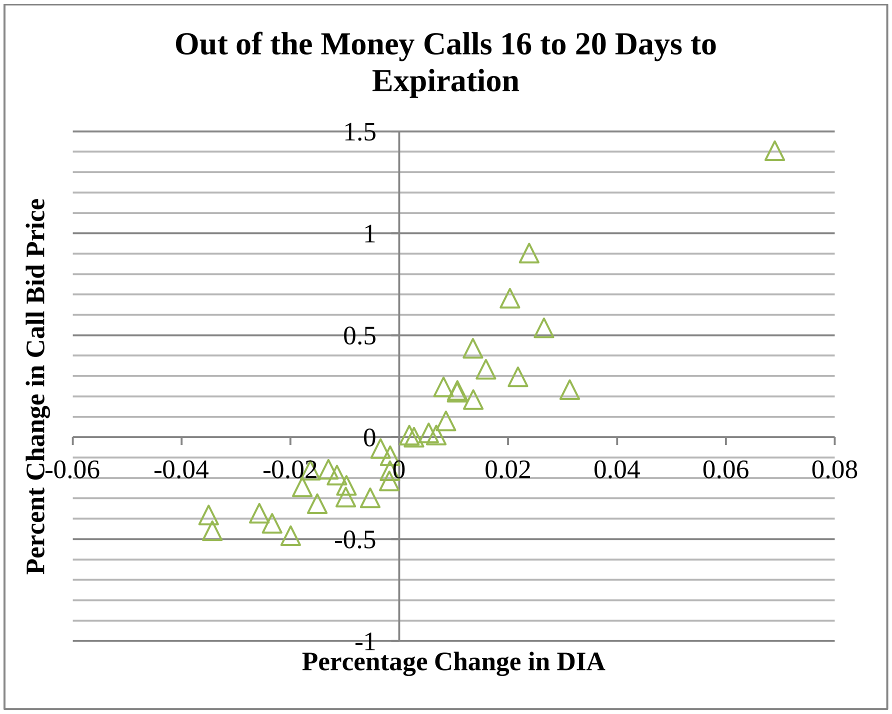

Figure 2. A plot of out-of-the-money calls 16 to 20 days from expiration. The triangles demonstrate the relationship between the percentage change in the DIA and the percentage change in the bid price of out-of-the-money call options 16 to 20 days from expiration.

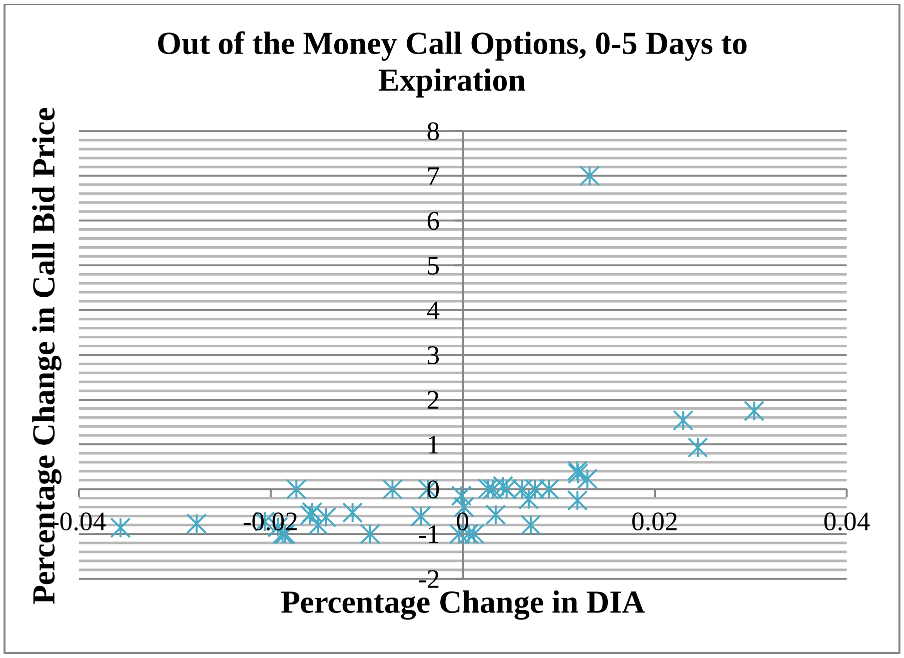

Figure 3. A plot of out-of-the-money calls 0 to 5 days from expiration. The symbols illustrate the relationship between the percentage change in the DIA and the Percentage change in the bid price of out-of-the-money call options, 0 to 5 days before expiration.

Appendix B Regression Results for Out-of-the-money Calls

Out of the Money	Coeff	Df	R-Squared	t-statistic	tdist	F statistic	Fdist
0 to 5	46.10	38	0.258	3.64	0.000	13.23	0.001
6 to 10	30.27	38	0.731	10.16	0.000	103.40	0.000
11 to 15	21.04	33	0.804	11.62	0.000	135.13	0.000
16 to 20	18.40	33	0.857	14.04	0.000	198.35	0.000