Spring 2014

Forecasting Short-Term Stock Returns Using Irregular Pricing Behavior in the Options Market

Thomas W. Sampson
thomaswsampson@hotmail.com

Forecasting Short-Term Stock Returns Using Irregular Pricing Behavior in the Options Market

by

Thomas W. Sampson

VITA AUCTORIS

Name:	Thomas W. Sampson
Date of Birth:	November 2, 1970
High School:	Estancia
Graduated:	June 1989
Baccalaureate Degree:	Economics
College:	University of California, San Diego
Date:	June 1993

Forecasting Short-Term Stock Returns Using Irregular Pricing Behavior in the Options Market

A Thesis presented to the Faculty of the Economics Department, University of San Francisco

---

in partial fulfillment of the requirements for the degree of Master of Arts in Financial Economics

---

by

Thomas W. Sampson

February 1, 2003

Forecasting Short-Term Stock Returns Using Irregular Pricing Behavior in the Options Market

Thomas W. Sampson

Tel: 415-703-0803
Email: twsampson@sbcglobl.net

February 1, 2003

ABSTRACT: This paper uses regression analysis to examine the relationship between today's implied volatility on AMD and OEX options with tomorrow's return on the underlying. An economic analysis of the options markets' microstructure is discussed to establish the intuition and the basis behind this relationship. Four separate models are developed to examine its statistical significance and the ability of options' prices to accurately forecast returns on the underlying security.

I find that today's call implied volatility has a significantly positive relationship with tomorrow's returns on AMD stock. Also, today's put implied volatility has a significantly negative relationship with tomorrow's returns on AMD stock. These statistical relationships are not evident between OEX returns and implied volatility on OEX options.

The hypothesis of the paper is that daily changes in implied volatility can be used to earn higher than expected returns on AMD stock. Two time-series models of AMD returns and actual sample returns generated by the forecasting models are used to test the validity and reliability of the models. I find that implied volatility can be used to increase forecasting accuracy and may provide a means by which the Efficient Markets Hypothesis can be refuted.

The main purpose of the paper is not to provide a method in which abnormal returns can be made possible, but to illustrate how it is possible for information to enter the options market before it enters the market for the underlying. Information contained in the options market microstructure, which can be quantified by analyzing implied volatility, may provide insight into the overall market's perception and expectation of future returns on the underlying.

Chapter 1: Introduction

From the Trading Floor to Academic Research

"Hey, Sampson! You know those calls you bought yesterday."
"Yeah?"
"News came out and the stock looks up two bucks. Your customer is a crook!"
"Sorry?"

I was a rookie options floor broker at the Pacific Exchange. The day prior to my above confrontation with a market-maker, I was buying Advanced Micro Devices (AMD) call option contracts for one of my customers. Throughout the day, the customer placed five separate orders to buy AMD November 30 calls. I purchased a total of 500 contracts. At the time AMD was at thirty dollars per share in quiet, low price activity trading. With each trade, the market-makers raised their offer on the November 30 calls from an initial price of two dollars up to 2 ½. Seeing that the price of the underlying stock remained constant, the rise in price was what the market-makers affectionately refer to as a "fish on-line"; a customer willing to trade on their market with no regard to price. I assumed that each order would be the last, because the price of the calls appeared to me to be exorbitantly high and thus the customer would be forced to back away. But he kept coming.

The next day AMD opened at 32 ¼. It was then apparent, as the verbal abuse I received from the market-maker made clear, that the customer must have known some kind of "inside information". This accounted for why he was willing to continue buying calls despite their "high" price. Confident that the stock would gap up the next day, the high premium was of little concern to him.

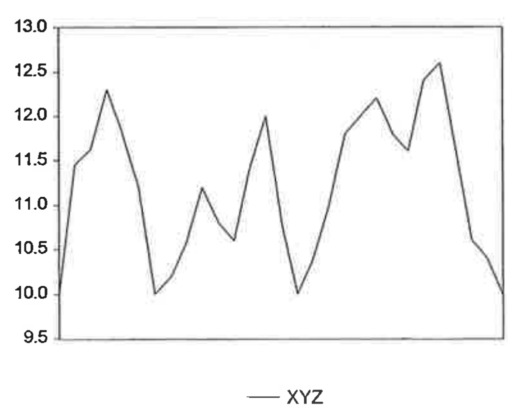
I relayed my "discussion" with the market-maker to my boss after the closing bell. He responded that I better become accustomed to that kind of verbal tirade because "our customers tend to be right". I took that as meaning that some of our customers were in fact crooks privileged to information about certain stocks that the options market-makers were not. During my three years on the options floor, I have seen several examples of off-floor traders taking advantage of unsuspecting on-floor traders. As a prospective proprietary trader myself, I became fascinated with these events. In this thesis I have set out to attempt to quantify these events and develop a statistical relationship between options and future movements in the underlying security. In so doing I hope to highlight the existence of a certain type of market inefficiency in the options market, that in turn might lead to a practical trading strategy.

Market Efficiency and Information

Simply stated, a market is efficient when past prices cannot be used to forecast future prices. On an intuitive level the Efficient Markets Hypothesis (EMH) seems quite logical and reasonable. If investors know that the price of a stock will go up next week, then they will undoubtedly buy the stock today pushing its price up to the future expected level immediately. Investing is an inherently selfish act where participants are intent on serving and increasing their own wealth. Competition among numerous traders allows financial markets to adjust quickly to any new information. The price of a security should therefore reflect all information at all times, making profitable forecasts not only difficult but also impossible.

Market efficiency implies that security prices follow a random path. Under EMH the best possible guess for tomorrow's price is today's closing price. The divergence in price from day to day is thus random and uncorrelated. However, historical stock market data suggests that financial markets do have a positive drift rate. Investors who have bought and held a diversified portfolio of stocks have reaped considerable rewards in the past decade. By stating the market is random, it is not academia's intention to assert that consistent, long-term profits are not possible, only that short-term profits based on some kind of forecasting model are. This is a difficult notion for many, especially high paid Wall Street executives, to accept. If it is impossible to achieve higher than expected profits by employing countless hours of technical analysis and research, then why should investors pay high commissions to brokerage houses for their services?

A technical analyst is a type of trader who disagrees with the theory of unforecastable changes. These traders, or "chartists", analyze patterns in historical price graphs to forecast stock prices. They would probably look at the following chart of XYZ and determine that when it trades down to ten dollars, the next day it bounces up providing a significant buying opportunity.



The financial economist would look at the graph and argue that the best guess for tomorrow's price is simply today's price of ten dollars. There is no reason to believe past behavior is indicative of future behavior. Even though such techniques receive considerable media coverage, they have little if any statistical credibility.

This discussion of market efficiency places us in an awkward situation. As a student of financial economics, I believe in the random nature of security prices and while the intent of my paper is to develop models designed to forecast short-term changes. The Efficient Markets Hypothesis is just that, a hypothesis. It must therefore be tested and continually challenged in order to prove its value and relevance. Given the widespread academic acceptance of the hypothesis it would be foolish of me to expect to find overwhelming evidence against it. The paper, in its examination of the derivatives market as a basis of the tests, does however provide a different and unique way of exploring market efficiency that may provide results that are in disagreement with traditional tests.

Traditional Market Efficiency Tests Vs. Derivatives Based Tests

Traditional tests of market efficiency range from the statistical time-series analysis of past prices to more complex models such as the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Model (APT). These tests focus on past prices of assets, the introduction of information to the market (i.e. news regarding dividends or stock splits), and the relationship between a stock and an index of stocks. This paper presents a different tool in testing EMH and thus a new tool in forecasting equity returns: options.

Options by definition are forward-looking. EMH implies a stock's price incorporates all past and current information while an option's price must also incorporate all possible future information and events associated with the stock until expiration. This fact, even with complex pricing formulas, makes accurate pricing of options an inexact science at best. The paper assumes and relies on certain situations in the market where the pricing of options incorporates information about future directional movement in the underlying asset, hereafter referred to as the "underlying". If these situations do occur with some regularity, the paper's models can be used in a statistically significant way to systematically predict returns on a stock. Even though the trader's information in my AMD story constitutes a violation of market efficiency, it would not be detected by traditional tests that focus on historical prices of the underlying only. The paper is forced to implicitly assume that the market is exhibits a certain type of inefficiency (the three types of market efficiency will be defined in Chapter 2), because it relies on the inside knowledge of others for its own accuracy and statistical strength.

Options Market Dynamics: Supply and Demand

The options trading-pits are one of the purest forms of free market capitalism. Buyers and sellers come together in an effort to maximize their own utility, ascertain a price agreeable to both parties, and enter into a binding, legal bargain. Derivatives are maybe the most complex and least understood financial product. Yet, the price of an option is dependent on the same principles of supply and demand and on the same kind of market-manipulating forces, as is the price of gasoline.

The options market is comprised of market-makers standing in trading-pits and off-floor traders who place buy and sell orders via a broker. The vast majority of trades occur between these two parties. Off-floor traders are dependent on the market-makers for not only the displayed quotes, but for market liquidity. In the AMD example there was an increase in demand for call options. The trading crowd, knowing they had an aggressive buyer of their options who made it clear he would pay any price for them, acted in their own self-interest. As experienced traders with considerable gamesmanship, they took full advantage of the situation by raising the price in order to enhance their own competitive position and in doing so were able to increase the supply to meet the increase in demand of the desired call option.

The market dynamics associated with AMD options are not universal. The OEX is an index option whose underlying security is a basket of stocks, the S&P 100. It is the largest options market in terms of volume and number of market-makers (approximately 200 versus only 10 in AMD). The driving force behind the change in market price for AMD November 30 calls does not exist to the same extent in the OEX. A purchase of 500 calls in the OEX would not significantly alter pricing because of the larger number of traders and thus greater overall liquidity. It is also safe to assume that competitive pressures make it difficult for a large group of self-interested traders to collectively agree to raise prices in order to bait a customer into paying too high of a premium. While the AMD options market demand/supply dynamics may allow for some forecasting accuracy, these same dynamics do not exist for the OEX.

These dynamics are simple in nature, but they are critical to the paper's assumptions and expectations. The paper assumes that these basic demand/supply dynamics in certain less heavily traded equity options markets do have an effect on prices. These changes in prices lead in turn to the expectations that they contain information about future returns on the underlying.

Calls and Puts: Hedging and Speculating

A call option gives the buyer the right, but not the obligation, to buy an asset for a prescribed price (the strike price) within a prescribed time period. The right to sell an asset is a put option and has payoffs opposite to those of a call. Whereas the holder of a call wants the price of the underlying asset to rise – the higher the asset price at expiration the greater the profit – the holder of a put option wants the asset price to fall. A rise in the price of the stock above the strike price gives the call holder the ability to buy stock below current market value. A fall in the price of stock below the strike price gives the put holder the ability to sell stock above market value.

Calls and puts serve investors needs in two ways: 1. Hedging 2. Speculating. As a hedging tool, options allow investors to protect their assets. The seller (or "writer") of a call immediately receives a premium, the market price of the option. This premium then reduces the downside the liability of owning a stock. For example, if an investor purchases stock for $50 per share and sells an equivalent number of calls with strike price of 55 for a premium of $2 per option, he will not incur losses as long as the stock stays above $48 per share. As with all things, protection does not come free. Selling calls limits upside potential. If the stock trades above the strike price (55), the stock will be "called" away forcing the investor to sell at $55 per share regardless of current market value. This strategy may be attractive to many traders, because if the stock is called away at a price far below current value, the trader still profits $7 per share. ($2 received for selling call + sell stock at $55 - $50 original price paid for stock = $7 profit)

Buying a put can limit stock losses to a specific amount. If I were to buy stock for $50 per share, I could hedge the position with the purchase of a put with a 50 strike price. This strategy guarantees me the right to sell stock at $50 regardless of how low the stock itself trades. The price of protection and the reduction of possible profit is the price of the put. Like car insurance, you never want to use it but it allows risk averse individuals to sleep easier.

Options also allow investors to speculate on short-term movements in a stock or index. Since the prices of calls and the underlying are positively correlated, the trader who believes a stock will rise can purchase a call now instead of the stock itself. The purchase of a put alone without holding underlying allows investors to profit from downward movements in the market. This strategy is popular among traders who try to "time" the market. Buying a call is less expensive than buying the underlying and buying a put generates a "short" position. The term "short" means the selling of an asset by an investor who does not already own the asset. This strategy is profitable when the price of the asset falls.

Expectations and Hypothesis

The introduction so far is intent on giving the reader enough background market knowledge to understand the intuition behind the paper's expectations and abilities to refute the EMH. Supply and demand affect the prices of all financial instruments regardless of how complicated their pricing structure is. In the case of options, these effects can be calculated by deriving implied volatility from the Black-Scholes pricing formula. An increase in demand for an option causes its implied volatility to rise while a decrease in supply causes it to fall. Today's implied volatility is then related to returns on the underlying on the same day and perhaps to returns on the underlying tomorrow.

Hypothesis:

1. Call implied volatility is negatively correlated with returns on the underlying when compared on the *same* trading day. Investors hedge the purchase of stock by selling calls. When there are more buyers than sellers at the current price, the price rises until the two are equated. More stock buyers mean more call sellers. Result: when the price of an asset rises the price of calls and implied volatility falls.

2. Put implied volatility is positively correlated with returns on a stock when compared on the *same* trading day. Investors hedge the purchase of stock by buying puts. Excessive simultaneous demand in both markets at the current price level causes the prices and put implied volatility to all rise.

3. *Today's* call implied volatility has a positive relationship with *tomorrow's* returns on a stock. High speculation today about future rises in a stock's price causes excessive demand in the calls raising their price and call implied volatility. The expectation is dependent on the accuracy of these speculators.

4. *Today's* put implied volatility has a negative relationship with *tomorrow's* returns on a stock. High speculation today about future falls in a stock's price causes excessive demand for puts raising their price and their implied volatility. As with the calls, if these speculators are correct in their forecast, a rise in put volatility today will result in negative returns on the underlying tomorrow.

5. The expected relationships hold only for an individual equity and not for an index due to options market dynamics.

The thesis attempts to provide a method to forecast movements in stock prices. I hypothesize that the prevalence of options traders who have inside knowledge regarding directional movement in the underlying is strong enough for Hypothesis 3 and 4 to hold. Due to the forward looking nature of options, this information can then be used to forecast stocks and earn higher than expected returns by removing a portion of the day-to-day randomness in stock prices. By modeling the relationship between implied volatility and stock returns in a particular way, signals to buy and sell the stock will emerge. A buy signal today will then result in significantly positive stock returns tomorrow. A sell signal today will result in significantly negative returns tomorrow. The models will also increase the overall statistical forecasting accuracy of the stock in comparison to time-series models.

The paper is structured as follows: Chapter 2 presents a more thorough discussion of the Market Efficiency Hypothesis, a discussion of the Black-Scholes pricing formula and how it is used to calculate implied volatility, interpretations of implied volatility's strength and weaknesses associated with its use as a tool to forecast stock returns, and summary statistics for all data. Chapter 3 details and gives the rationale for the regression models and presents the statistical significance of the forecasting models. It also presents evidence for the forecasting accuracy and the ability of the models to be used as a successful trading strategy. Chapter 4 contains concluding comments.

Chapter 2: The Details

The financial world, once exclusive to, and dominated by a select group of firms and wealthy individuals, is now open to a wide variety of investors. These investors range from the traditional and powerful Wall Street institutional traders to middle income families who trade stocks from home through the Internet. The common thread that ties different investors together is a belief, or an assumption that they have the tools and information necessary to profit from their trading strategies. Our society has become fascinated with the countless riches that it seems possible to make in the stock market. A quick glance at the finance section in a bookstore will show just how popular this new "national pastime" has become. Thousands of books and articles, written by self-described "investment professionals", tout sure-fire ways to "beat the market". Methods range from complex (yet, easy to understand if you read their book) charting techniques to advanced neural network computer programs that assimilate vast amounts of data to produce market forecasts.

One of the most pressing, and fascinating, questions facing financial economists is whether or not one or more of these many trading strategies can be used to predict market movements and earn higher than expected profits. The underlying assumption or null hypothesis is that it is not possible and the stock market is efficient. The logic behind this assertion is that any price discrepancy that allows for risk-less profit will be acted upon until it no longer exists.

In this chapter I will expand on the previous chapter's discussion of market efficiency and also explore possible causes of inefficiencies and their relation to the many predominant "beat the street" trading techniques. Implied volatility is defined along with its strengths and weaknesses as a means to forecast market returns. A brief discussion of the Black-Scholes options pricing model and how it is used to derive implied volatility is also presented. Chapter 2 contains a statistical summary and the method used for calculating the data sets contained in the forecasting models that follow in Chapter 3.

Three Types of Market Efficiency

The Efficient Markets Hypothesis is divided into three categories: Weak, Semi-Strong, and Strong Form Efficiency.

Weak Form Market Efficiency Hypothesis

Weak form efficiency states that past prices and stock movements cannot be used to predict future prices. It is a direct attack on technical analysis. According to this hypothesis, changes in a stock's price from day to day are random. Wall Street traders, who spend countless hours staring at intricate graphs of past stock prices, looking for trends and patterns, are simply wasting their time and the money of investors who pay for their advice. Random daily stock price changes lead to the most basic of conclusions regarding future prices: the best guess for tomorrow's stock price is today's closing price. Any and all variation is random and thus not predictable. This is termed a random walk in the stock's price.

Weak efficiency is thus statistically defined by "The Random Walk". Econometric analysis of daily returns on a stock yield three types of random walks that are all in agreement with the Weak Form Efficiency Hypothesis.

The Random Walk Hypothesis and Weak Form Market Efficiency

"Random walk" is a term used to describe the behavior of a certain type of time-series random variable. It also captures the behavior of stock prices that conform to the Weak Form Efficiency Hypothesis. The first property of a random walk is the property that the best guess for tomorrow's price, given information available today, is today's price or $E[P_{t+1} \mid I^{w}_{t}] = P_t$. Where I^{w}_{t} = "information" regarding the price of the stock and in the case of weak form efficiency the information that is of concern only involves past prices and returns (i.e. "charts"). $E[P_{t+1} \mid I^{w}_{t}] = P_t$ is then defined as tomorrow's expected stock price *given* today's information about past stock prices is equal to today's stock price. Since the weak form does not disagree with the notion of a market that can drift in a long-term direction, this expected price change can be included in the equation as μ_t. The last component is the error term (ε_{t+1}). The error term is also unforecastable under weak form efficiency or $E[\varepsilon_{t+1} \mid I_t] = 0$. The dynamics of a random walk with a drift in a stock's price P_{t+1} are thus given by the following equation:

$$P_{t+1} = P_t + \mu_t + \varepsilon_t$$

By the definition of weak efficiency, the error term, or daily, unexpected change in price, should be random. The Random Walk Hypothesis breaks down the statistical nature of the error term into three types.

Random Walk 1 (RW1) entails an error term that is independently and identically distributed (IID). IID is equivalent to complete randomness, or white noise, in that the error term is normally distributed with mean of zero and a constant variance (θ^2). White noise error terms by definition cannot be forecasted with available information and thus a stock with this property is not completely predictable using past prices. If the error terms are IID and normally distributed, then there is a positive probability that P_{t+1} will be less than zero. It is obviously not possible for a stock's price to be negative. This is why this thesis, and financial research of the same type, uses the lognormal of stock price time-series data.

In Random Walk 2 (RW2) the assumption in RW1 that the daily, unexpected change in price is IID is relaxed. RW2 is probably more realistic for the case of financial data. The notion that the variability of stock returns is constant over long periods is not realistic. Error terms that are independent but not identically distributed (INID) allow for the possibility of unconditional heteroskedasticity in the time-series. This is a particularly useful feature when dealing with stock returns. When viewing the daily volatility of a stock's price as a scatter diagram, it often appears that some periods exhibit greater volatility than others do. This unequal scatter (heteroskedasticity) allows for differing volatility while maintaining the property that ε_{t+1} is unforecastable with I_t.

Random Walk 3 (RW3) is the most general of the three versions. RW3 relaxes the assumptions on RW2 by allowing for pricing data not to be independent. Daily returns can therefore have dependent but uncorrelated error terms. Under RW3 the data is uncorrelated in that the covariance between ε_t and ε_{t-k} is zero for all k but the covariance between squared error terms (ε_t^2 and ε_{t-k}^2) is not independent in that their covariance does not equal zero. As we will see in the discussion of the Black-Scholes options pricing formula, the ability to forecast volatility is a vital component of accurately priced derivative securities. Under the Weak Form Efficiency Hypothesis, even though volatility may be predictable to a certain degree, actual prices are still not predictable.

Semi-Strong Form Market Efficiency Hypothesis

Semi-strong efficiency builds on the weak form by allowing all published information to be used in forecasting future price changes. In the equation describing the dynamics of Pt+1, I^{w}_{t} is replaced by I^{ss}_{t}. I^{ss}_{t} is now defined as all information regarding past prices and all published information about the stock. Fundamental analysts are traders that perform research on, and publish information about, the inner workings of a publicly held company. Traders then use this public information, ranging from price-earnings ratios to product inventory levels, to make investment decisions. Semi-strong efficiency claims that the price of a security reflects this information at all times and thus it cannot be used to forecast short-term price fluctuations. No matter how hard you try to analyze past price performance along with *all* publicly available news and information about a company, you will be unable to forecast future returns if the market is semi-strong efficient.

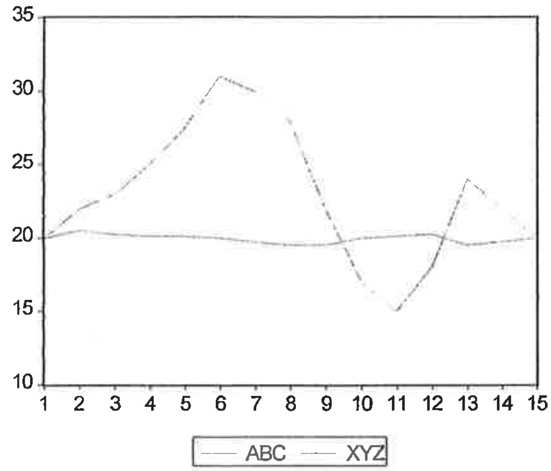
Strong Form Market Efficiency Hypothesis

The strong view of market efficiency is the most direct: nothing – not even unpublished ("inside") information – can be used to predict stock returns. The restrictive nature of this hypothesis makes it the easiest to refute, which is normally the case in studies pertaining to the subject. There are an abundance of stories about people on the "inside" of a company who use their (private) information to generate trading profits. For example, a corporate lawyer who is hired to negotiate a merger between two companies may be able to use this information to predict the movements in the associated stocks. A government agency, the SEC, was established in part to punish and attempt to prevent such illegal trading activity.

Implied Volatility and the Black-Scholes Options Pricing Formula

The Black-Scholes formula (1973) revolutionized the financial industry by providing an easy-to-use model that adequately prices derivative securities. The model assumes stock prices follow a lognormal random walk (geometric Brownian motion). It has been shown that stock prices exhibit more complicated behavior which has led to more complicated pricing formulas. The ease in which Black-Scholes can be applied to real-time trading situations, however, more than makes up for its less-than-perfect description of asset behavior. The ease of Black-Scholes can be seen by the fact that only one parameter, future volatility, is not directly observable in the market. The other parameters (stock price, time to expiration and the interest rate) are always known with certainty.

We know that an asset's volatility tends to exhibit heteroskedasticity. From the RW3 discussion we also know that it may be possible to infer with some accuracy what tomorrow's volatility will be based on today's volatility. The ability to adequately make these types of assertions regarding daily changes in volatility allows traders to implement the Black-Scholes model even though actual future volatility is not known with certainty. This does not undermine the overall importance of volatility in the accurate pricing of an option. The following graph of stock prices is presented to illustrate this point.



Both stocks ABC and XYZ have an initial price and a price at expiration (day 15) of $20. But a call option with strike price of 25 will trade for a higher price on stock XYZ than on ABC. Assuming all other parameters constant, the reason for this is volatility or variance in the price of the two stocks. From the graphs, it is visually obvious that XYZ has a greater variance and thus greater probability of trading in-the-money (above $25). If the correct volatility is not used to price the options in the marketplace, investors could trade options and earn a higher than expected return.

This discussion leads back to the principles of market efficiency. If the market is in fact efficient, then the volatility parameter used in the options market should not only contain information about current changes but also information about future changes in the stock's price. It is already assumed then that the volatility used to price an option gives information regarding the *size* of future changes. It does not necessarily mean, however, it gives information regarding the *direction* of these changes. Thus it may not contain information useable in the forecasting of returns.

Implied Volatility

As noted above, the one parameter in the Black-Scholes model that is not directly observable is volatility. Volatility can be estimated using historical trends in the stock price. An alternative approach is implied volatility or the volatility of the underlying "implied" by an option price.

Since all other Black-Scholes parameters are known with certainty, along with the price of an option, through an iterated process the correct volatility can be found within the Black-Scholes pricing model's framework such that the resulting calculated option price equals the option price found in the marketplace. This process allows for the implied volatility to reflect not only the market's opinion of actual and future volatility but since it is dependent on the price of the option it is also dependent on, and therefore must also reflect, outside forces affecting the options' pricing structure. These forces are demand/supply market dynamics or the "microstructure" of the market. To illustrate these points assume a stock's price remains constant along with all observable parameters. As was the case in my AMD experience, a large trade occurring in the options market affects its pricing structure causing implied volatility to change. This change is not due to changes in stock volatility but rather to reactions to demand or supply (microstructure inefficiencies). These are the types of changes that must be quantified in order to achieve successful forecasting models. The main advantage in using implied volatility as a forecasting tool is its implicit forward-looking nature.

Implied volatility is a parameter defined to indicate future dispersion in price. The goal of the models is then to find events in the data that provide not only information about the size of future movements, but also about the direction of these movements.

The Data

The forecasting models use three types of time-series data: daily returns on the underlying security, call implied volatility and put implied volatility in the corresponding option class.

The stock and equity index data to be analyzed is AMD and OEX respectively. AMD or Advanced Micro Devices, is a Silicon Valley based company that produces computer chips. AMD is currently Intel's main rival in the chip industry. AMD stock trades on the New York Stock Exchange while its options trade on the Pacific Exchange. "OEX" is the option ticker symbol for the Standard and Poor's 100 index. The index is comprised of one hundred of the largest capitalized companies in the United States. It is intended to be representative of all major industries and thus its movement should be indicative of the strength or weakness in the entire market. OEX options trade on the Chicago Board Options Exchange.

As discussed earlier, daily changes in the securities are calculated using natural logarithms. The notations used for daily natural logarithmic returns are R(AMD) and R(OEX). The following formula is used for their calculation:

$$R(AMD)_t = \ln (AMD_t / AMD_{t-1})$$
$$R(OEX)_t = \ln (OEX_t / OEX_{t-1})$$

Stock splits and dividends complicate any calculation of daily returns on financial time-series. During the course of the sample period (one year or 252 trading days) AMD did not incur a split in price nor a dividend. The OEX also did not split but since it is an index it has what is known as a "dividend-yield". This is as annualized number that incorporates the return associated from paid dividends on all hundred stocks.

Dividends have the effect of reducing the stock price on the ex-dividend date. For tax reasons the reduction is not one-for-one. It is generally assumed for statistical analysis that if a stock goes ex-dividend for $1 per share, the stock's price will be reduced by $0.80 or 80% of the dividend amount. To avoid problems associated with continually adjusting the index price in accordance with paid dividends, the OEX changes a stocks divisor when it goes ex-dividend in order to keep the index value unchanged. The "divisor" is the percentage of the total index value that a single stock accounts for. By reducing this number for an ex-dividend stock (and raising the percentage composition of the other stocks), the index price remains constant. This continual, yet very minor, tinkering of the index allows for the above formula to be an accurate calculation of daily returns.

The call and put implied volatility data is calculated based on the daily closing option price for each series. Each individual option price is then fixed in the Black-Scholes pricing model and through the process of iteration, each option's implied volatility is determined. The implied volatility number is an annualized percentage based on 252 trading days in a year. If a stock is said to have a price volatility of 50, it equates to the maximum dispersion of price over one year to be fifty percent of the current price in both directions. For example, the annual price range for a ten-dollar stock with a fifty volatility is *expected* to be between five and fifteen dollars.

In order to generate a single call and put implied volatility for each trading day, a weighted-average system based on trading volume is used. The weighting factor is the volume of each call or put series divided by the total number of call or puts traded during a single day. The weighting factor is then multiplied by each option's implied volatility. The sum of all calls and puts volume weighted implied volatility is used as Call IV and Put IV. The daily change in implied volatility is calculated in the same manner as the daily returns on the underlying.

$$\text{Call IV}_t = \sum (\text{weighting factor}_{i,t} \times \text{Call IV}_{i,t})$$
$$R(\text{Call IV})_t = \ln (\text{Call IV}_t / \text{Call IV}_{t-1})$$

Since at-the-money options (options whose strike price is close to the actual price on the underlying) generally have larger trading volumes than deep in-the-money or out of–the-money options, this method allows for them to carry more weight in the total calculation of implied volatility. This fact is important since the pricing of at-the-money options are more sensitive to changes in supply and demand. Deep in-the-money options move tick for tick with the stock. Their time premium is limited to the current interest rate (i.e. the time value or opportunity cost associated with using dollars to purchase stock) and thus their price is not affected by demand/supply dynamics in the options market. Out of-the-money options have minimal price movements due to their small value and delta and thus their pricing structure does not exhibit significant changes due to demand/supply dynamics. (Delta is defined as the change in an options price per change in the stock's price. Deep in-the-money options have a delta close to 100 or their movements are 100% the movement of the stock. Out of-the-money options may have a delta around 5. If a stock's price rises one dollar, the out of-the-money call option will only rise in value by five cents.) The change in pricing due to trading patterns is the focus of this study and since the price of at-the-money options have a greater chance of being significantly affected by these situations, this type of weighting structure is the most relevant and useful.

Summary statistics for daily returns, call and put implied volatility:

AMD *Summary Statistics*	Mean	Std. Dev.	Skew	Kurt	Max	Min	Jarque-Bera	Norm. Prob.
AMD	31.55	9.062	-0.027	1.515	47.375	17.125	na	na
R(AMD)	-0.0025	0.0449	-0.1019	7.284	0.1744	-0.2208	192.4	0%
AMD Call IV	62.71	11.29	0.487	2.972	105.25	38.4	na	na
R(AMD Call IV)	0.0006	0.1261	-0.0521	7.043	0.5546	-0.5371	175.2	0%
AMD Put IV	62.58	11.23	0.533	2.818	101.41	43.28	na	na
R(AMD Put IV)	-0.0008	0.1129	0.0447	6.681	0.5319	-0.4249	149.9	0%

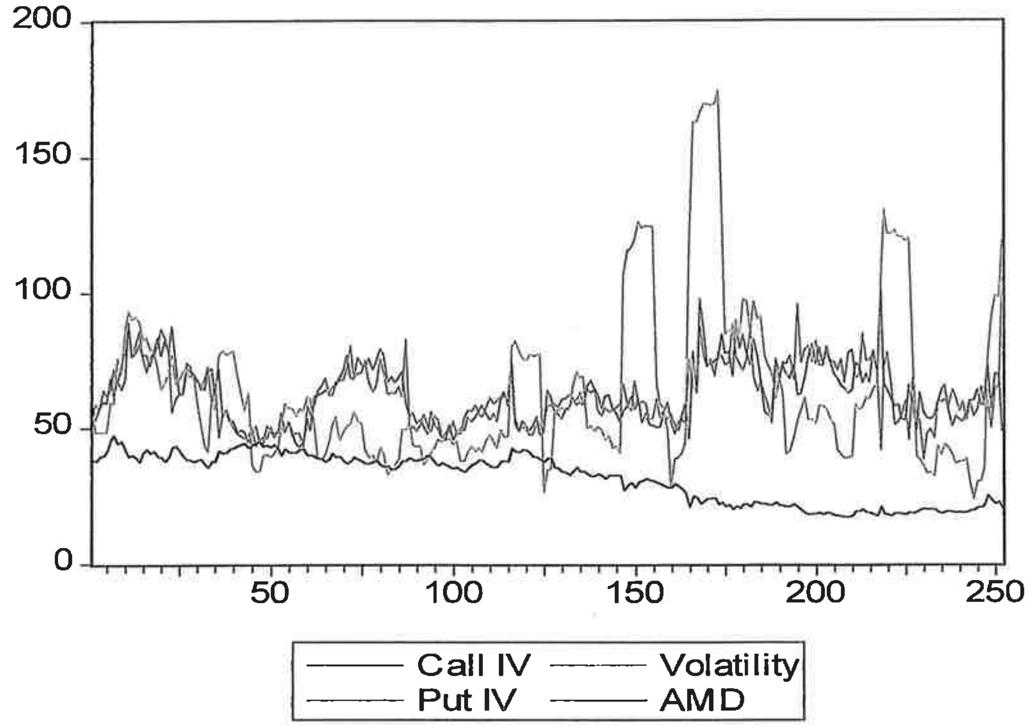
OEX *Summary Statistics*	Mean	Std. Dev.	Skew	Kurt	Max	Min	Jarque-Bera	Norm. Prob.
OEX	487.69	40.08	0.472	1.743	576.24	417.6	na	na
R(OEX)	0.0015	0.0091	-0.323	3.299	0.0214	-0.0301	3.03	22%
OEX Call IV	19.519	2.579	0.788	3.681	27.02	14.5	na	na
R(OEX Call IV)	-0.0023	0.0725	0.459	3.658	0.2263	-0.1679	7.66	2%
OEX Put IV	19.846	2.744	0.793	5.282	29.71	10.09	na	na
R(OEX Put IV)	-0.0017	0.1052	-0.333	14.57	0.512	-0.584	805	0%

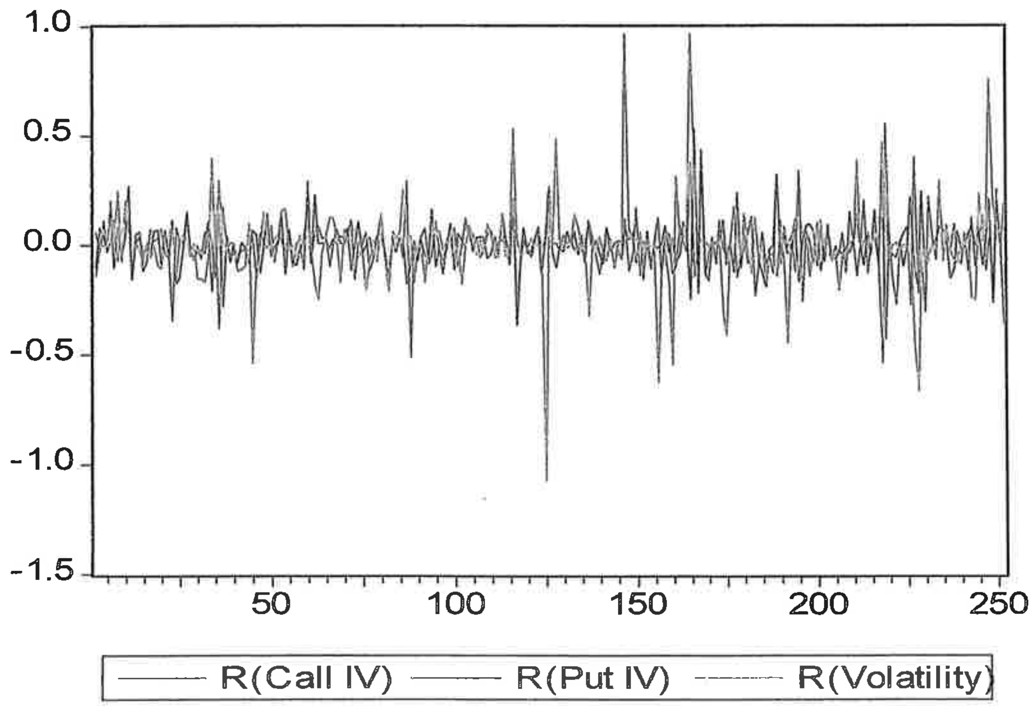
From the Jarque-Bera statistic it can be seen that the returns on the two data sets have a 0% probability of being normally distributed. The fact that the returns do not exhibit a normal distribution is evidence of an inherent flaw in the Black-Scholes pricing formula. The paper is dependent on this model for its calculations of implied volatility and thus it may incur a similar weakness or bias in its forecasting models. However, since a vast majority of professional traders use and have confidence in the Black-Scholes models, it will be assumed here that any flaws are small enough as to not significantly affect any possible conclusions.

A quick glance at the tables will also show a dramatic difference in the mean and standard deviation of AMD implied volatility and that of the OEX. A single stock, especially a stock involved in high-tech products is more likely to exhibit higher and more dramatic changes in volatility than that of a diversified index of stocks. This too may account for why the models work for AMD and not for the OEX. The models are dependent on sudden and large shifts in implied volatility for its trading signals. If these events do not occur with regularity, as is the apparent case with the OEX, the models are less valid if not worthless.

The graphs shown on the following page are presented as "eyeball" evidence for the relationship between AMD's stock price, implied volatility and actual AMD price volatility. The first graph is actual daily values. As expected, it shows how implied volatility mirrors movements in actual volatility. Implied volatility appears almost to be a moving-average of volatility, following the same path yet unable to account for sudden and dramatic changes. This is in fact how the option price parameter is routinely calculated. The second graph illustrates the daily logarithmic changes in the volatility data.

The "eyeball" evidence to be noticed in the second graph is the occurrences where changes in call and put implied volatility diverge from each other. Under the Black-Scholes pricing model the volatility parameter is and should be the same for both call and put options. This is generally true but then why do the two implied volatility parameters differ significantly on certain trading days? The paper hypothesizes that this difference is caused by changes in demand/supply dynamics of the two option series.





The following chapter will define and explain the forecasting models and how they attempt to quantify specific changes in implied volatility, such as a significant difference in call and put implied volatility, and the intuition behind why certain changes may lead to stock price forecasting accuracy.

Chapter 3: The Models

Chapter 3 presents four different "Forecasting Models". These linear regression models are developed with the intent of finding a statistically significant relationship between today's call and/or put implied volatility with tomorrow's change in the underlying price. The existence of such a relationship could be used to earn higher than expected returns and provide evidence against the Weak Form Market Efficiency Hypothesis.

As current or potential traders, we are not necessarily concerned with significant coefficients or goodness-of-fit values. True interest and value lies with a model that explicitly tells us when to buy or sell a security. So in addition to the statistical data from the regression models, buy and sell "signals" are developed. These signals are then used to examine the "real world" outcome of the trading strategies if they were employed during the sample period. The thesis hypothesizes that the AMD forecasting models not only have significant estimated regression coefficients, but the next day returns on AMD are significantly different from expected returns and in agreement with expectations: a buy signal today results in positive returns tomorrow while a sell signal today results in a decrease in the price of AMD stock tomorrow. The thesis also hypothesizes that because these results depend on microstructure inefficiencies in the market they are less likely to be evident for the case of the OEX.

Chapter 3 is divided into two sections. The first section looks at the relationship between call/put implied volatility and returns on the underlying for the same trading day. Correlation matrices are used to give evidence for the assumptions regarding the role of demand/supply dynamics in the AMD options market and also how these assumptions are not evident in the OEX market. The second section details the four regression models used to forecast returns. These "Forecasting Models" relate today's call/put implied volatility to tomorrow's return on the underlying. Corresponding trading strategies are then developed along with an evaluation of their statistical significance and of their ability to accurately forecast next day returns.

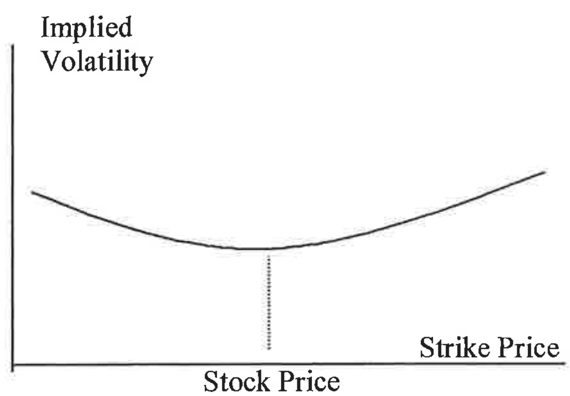
Same Day Relationship

As outlined in Chapter 1, it is expected that changes in call and put implied volatility have a statistical relationship with changes in the price of an equity for the same trading day. These changes in implied volatility are caused by demand/supply dynamics and are independent of changes in the options' price caused by other Black-Scholes price determinant variables. This relationship and thus this market microstructure are not the same for the case of options on the OEX.

Options are a tool used by investor to hedge positions in the underlying. When investors buy stock they can hedge their purchase by either selling calls or buying puts. From the discussion of demand/supply dynamics, it follows that an increase in the number of call sellers, resulting from an increase in the number of buyers of the underlying, reduces the price of call options independent of other factors and thus a corresponding decrease in call implied volatility will occur. For the case of puts, an increase in buyers will increase the implied volatility when a corresponding increase in demand for the stock exists. An opposite movement in the implied volatility of calls and puts is evident when the price of the stock decreases.

The price of calls is positively correlated with the price of the underlying. However, the implied volatility of call options on AMD is negatively correlated with the price of AMD stock during the same trading day. This is an important distinction to be made. It is possible for the price of the option to increase while its implied volatility. This is because there are two forces at work here: the Black-Scholes variables that affect the price of the option and the specific, irregular occurrence of microstructure market inefficiency in AMD options that affect the implied volatility.

The notion of a *volatility smile* can be used to gain insight into the assumptions regarding the role of demand/supply dynamics in affecting implied volatility. As shown below, a volatility smile is a plot of implied volatility as a function of its strike price.



The smile has been developed by traders as a way to account for imperfections in the Black-Scholes model. In-the-money and out of-the-money options tend to exhibit greater implied volatility than do at-the-money options.

From the volatility smile, as the price of the stock increases, more call option series move in-the-money and thus their implied volatility would tend to increase. But this result is counter to the demand/supply dynamics assumption regarding changes in implied volatility in relation to changes in the price of a stock during the same trading day. However, this fact does not counter the thesis' assumptions. It can be viewed as evidence for the strength of demand/supply dynamics in the options when the data is viewed on a daily basis.

If we assume the volatility smile to exist, then the demand/supply forces that affect the same day relationship must be strong enough to counter the relationship defined by the volatility smile. These microstructure dynamics may only occur for a brief period of time (i.e. one trading day) and then over the long-term (i.e. until the option expires) the relationship can be defined by the smile. Shown below is the correlation between returns on the underlying and call/put implied volatility. The implied volatility data is calculated in the same manner as it is in the forecasting models. A complete description of the data is in the following section.

Correlation								
	Implied Volatility Data							
	Call IV_t	Put IV_t	C $H\text{-}IV_t$	P $H\text{-}IV_t$	$C > P_t$	$P > C_t$	C AR_t	P AR_t
$R(AMD)_t$	-0.179	0.103	-0.261	0.217	-0.175	0.353	-0.323	0.245
$R(OEX)_t$	-0.065	-0.012	-0.087	-0.002	0.001	-0.015	-0.091	0.002

Forecasting Models

The forecasting models analyze the relationship between today's option implied volatility and tomorrow's return on the underlying. Four separate regression models are presented to give evidence for and to define the nature of this statistical relationship which will then be used to forecast returns on the underlying.

The models' assumptions are generated based on the idea that options can be used as a speculative tool. An investor with "inside" knowledge regarding a stock and thus knowledge about near term movements in the price of the stock can utilize and profit from this information by buying or selling options today. If an investor *knows* that a stock's price will rise tomorrow, he can purchase calls or sell puts today. If the increase in demand for calls and/or supply of puts if great enough to affect their pricing structure, a model may be developed to quantify these types of changes caused by the market's microstructure. In the regression analysis today's call and put implied volatility are the independent variables and tomorrow's return on the underlying is the dependent variable.

The estimated coefficient for the call implied volatility variables is expected to be significantly greater than zero. An increase in call demand can be viewed as information entering the market regarding future changes in the price of the underlying. If this information does not enter the actual market for the stock, rather only the options market, then the options' price reflects information not contained in the stock's price. Under the Efficient Markets Hypothesis, a stock's price incorporates all information at all times. But if we can extract information about the stock's price that is not currently incorporated into it, an inefficiency may be found. For the case of call implied volatility, a positive coefficient implies that the stock is currently undervalued and once the information contained in the options' price enters the stock market, the price of the stock will increase.

The same logic also implies to the relationship of today's put implied volatility with tomorrow's returns on the underlying. If information enters the options market and not the stock market through an increase in demand for puts, this may lead to the possibility that the stock is currently overpriced. The absence of this information in the stock market may lead to a decrease in the price of the stock once the "new" information is revealed. A corresponding increase in put implied volatility might then allow for a forecast of future decreases in the price of the stock. A significantly negative coefficient on the put implied volatility independent variable quantifies this relationship and is thus expected in the subsequent regression models.

The hypothesis regarding the statistical significance of the implied volatility coefficients is in accordance with the speculative and forward-looking nature of options. It also illustrates how the thesis is force to assume not only a violation in the Strong Form Efficient Markets Hypothesis, but in the widespread occurrence of such illegal trading activity. The forecasting models attempt to capitalize on information that is not currently incorporated in the price of the stock but is incorporated in the price of the options. The following four regression models attempt to capture and quantify market information by modeling implied volatility in such a way that the independent variable data reflects changes in the demand/supply dynamics of the options market.

Forecasting Model 1: "Simple"

$$\mathbf{R(Underlying)_t} = \alpha + \beta_1\, \mathbf{R(Call\ IV)_{t\text{-}1}} + \beta_2\, \mathbf{R(Put\ IV)_{t\text{-}1}} + \varepsilon_t$$

The Simple Model is just that. It contains daily changes in the data for the entire sample period. Tomorrow's (day t) returns on the underlying is regressed against today's (day t-1) change in call and put implied volatility. Since all the data is incorporated and there is no attempt to find *irregular* periods of implied volatility caused by market microstructure, the model is expected to and is the weakest in terms of statistical significance. However, the estimated coefficients (β_1 and β_2) are opposite in sign to what the correlation is for call/put implied volatility and returns on the same day. The fact that β_1 is positive and β_2 is negative (though not statistically significant) in the Simple Model is an indication that the assumptions regarding the information contained in options, when used as a speculative tool, may be justified. These preliminary results lead to the idea that if the independent variables can be correctly specified, the signs on their estimated coefficients will not only be in line with expectations but will be significant which then may lead to forecasting accuracy.

Forecasting Model 2: "High-Implied Volatility"

$$\mathbf{R(Underlying)_t = \alpha + \beta_1 \, R(Call\ H\text{-}IV)_{t-1} + \beta_2 \, R(Put\ H\text{-}IV)_{t-1} + \varepsilon_t}$$

The High-Implied Volatility Model allows for an analysis of the relationship between tomorrow's returns on the underlying and only with specific episodes of high-implied volatility today. "High" is defined as a trading day where the change in implied volatility is greater than one standard deviation above its sample mean. A dummy variable regression model is used so that the defined events retain their value in the estimation process while all other events are assigned a value of zero and thus have no impact on the calculations.

It is expected that the estimated coefficients are in accordance with the hypothesis (positive for call IV and negative for put IV) and also, more importantly, their absolute values should be greater than the corresponding coefficients in the Simple Model. If the pricing of the options is affected to a greater degree (i.e. greater daily change) then so should the price of the underlying change by a greater amount. The weakness in the model lies in the fact that these larger changes in implied volatility may not necessarily be caused by market microstructure. It may be more reasonable to assume that the high-implied volatility changes are caused by an increase in actual volatility in the price of the underlying. Implied volatility by definition is an estimate of volatility. Therefore if stock price volatility is high, so will be implied volatility. Even though the results are in line with expectations, since increased actual volatility only indicates greater price dispersion (up or down) and not a one directional change, the forecasting results may and should be weaker and less reliable than the subsequent models.

Forecasting Model 3: "Call vs. Put"

$$\textbf{R(Underlying)}_t = \alpha + \beta_1\ \textbf{R(Call H-IV vs. Put)}_{t-1} + \beta_2\ \textbf{R(Put H-IV vs. Call)}_{t-1} + \varepsilon_t$$

The weakness in the previous model leads to the intuition behind the Call vs. Put Model. By examining call and put implied volatility in terms relative to each other, it is a safer assumption that the model is uncovering changes in options prices caused by demand/supply dynamics and not by the volatility in the underlying.

By definition the implied volatility variable in the Black-Scholes model is the same for both calls and puts in most circumstances.[1] Volatility affects the price of calls and puts in the same way and thus there is only one estimation of volatility (implied volatility) which is thus used for the pricing models of both calls and puts. From this definition it follows that if the volatility estimation is different in the two types of options

---

[1] A case where puts are priced higher than calls is when the underlying is difficult to borrow. If an investor is unable to "short" a stock, the same position can be obtained by buying a put. Since the market-makers are in a similar position, they are less willing to sell puts since they cannot hedge the position through a short sell of the underlying. The net result is the price of puts is raised reflecting a reluctance of the market makers to sell them. This price disparity between calls and puts should therefore not be inferred as an occurrence in line with the model's definition of the independent variables. Since AMD stock during the sample period was "easy to borrow", this circumstance did not occur and all pricing discrepancies are caused by the model's overall assumptions

it cannot reflect a difference in opinion about future volatility. It must therefore reflect a force outside the pricing model's variables that is affecting the market price of the options. This force is assumed here to be the market microstructure. For example, simple demand/supply dynamics dictates if an abundance of call buyers and put sellers enter the market, the price of calls will go up and the price of puts will go down. This fact will have the same result on implied volatility since all other pricing parameters are the same for both calls and puts and as mentioned earlier are always known with certainty. Since market microstructure can be the only cause for dispersion in call and put implied volatility, this model is the best hope for quantifying such demand/supply affects on the options market's pricing structure. It is thus the best hope for uncovering information from the options market that is not currently incorporated in the price of the stock in its separate market.

The development of the model's independent variables requires calculating the daily difference between R(Call IV) and R(Put IV). The sample mean and standard deviation of this new time-series is then calculated. "Call H-IV vs. Put" is defined as a trading day where R(Call IV) minus R(Put IV) is greater than one standard deviation ***above*** its sample mean. The change in call implied volatility on the trading day in which the defined event occurred is then used in the regression using the dummy variable technique. Similarly, "Put H-IV vs. Call" identifies episodes where the difference is greater than one standard deviation ***below*** its sample mean. Once again, only the actual put implied volatility change for the trading day is used when its disparity with call implied volatility is in accordance with the above definition. Below is a table summarizing the statistics for the call and put implied volatility data incorporated in Model 3.

AMD *Model 3 Data*		Model 3 Independent Variables *Summary Statistics for R(Call IV) and R(Put IV) when Model 3 defined event occurs in sample*	
	R(Call IV) - R(Put IV)	**R(Call IV)**	**R(Put IV)**
Mean	0.0032	0.1713	0.1243
Max	0.9796	0.5546	0.5319
Min	-0.8792	-0.0326	-0.0123
Std. Dev.	0.1872	0.1321	0.1517
Count	252	25	18

Forecasting Model 4: "AR Shocks"

$$\mathbf{R(Underlying)_t = \alpha + \beta_1\, R(AR\ Call\ H\text{-}IV)_{t-1} + \beta_2\, R(AR\ Put\ H\text{-}IV)_{t-1} + \varepsilon_t}$$

The AR Shocks Model assumes the two implied volatility time-series follow a first-order autoregressive process or $R(IV)_t = \alpha + \rho\, R(IV)_{t-1} + \varepsilon_t$. The parameter ρ is the first-order serial correlation coefficient. Since the paper focuses on the importance of daily changes in implied volatility, this type of time-series estimation is a logical approach. Also, from the discussion on random walks, simple time-series forecasting techniques have a greater chance of accuracy when dealing with volatility of prices when compared to the forecasts of actual prices. The following table shows that the AR(1) estimation results are statistically significant.

	AMD Call IV	AMD Put IV	OEX Call IV	OEX Put IV
ρ	-0.405	-0.403	-0.151	-0.338
Std. Error	0.058	0.059	0.083	0.079
t-stat	-6.889	-6.783	-1.811	-4.274
Prob. $\rho = 0$	0%	0%	7.23%	0%

While this estimation is not intended or assumed to be the best model to forecast implied volatility, the results appear to be significant enough for the purpose of

Forecasting Model 4. The results of the model appear to show that high-implied volatility today predicts low-implied volatility tomorrow (ρ is significantly less than zero at the %1 level for all time-series accept OEX Call IV).

Forecasting Model 4 uses the residuals from the AR(1) estimation process. The residuals are viewed as "shocks" or unexpected changes in implied volatility. But what if there are shocks large enough such that they were the result of changes in the options demand/supply dynamics and not the result of changes in actual volatility in the underlying? These are the types of daily changes in implied volatility that the model attempts to uncover. "AR H-IV" is defined as a daily change in implied volatility where the residuals from the corresponding AR(1) estimation are greater than one standard deviation above their sample mean. The dummy variable technique is used once again so that only specific R(Call/Put IV) data that corresponds to a day where the model's defined event occurs is used in the regression. Below is a table summarizing the statistics for the AR(1) time-series estimation residuals and the implied volatility data used in Model 4.

AMD *Model 4 Data*	**Call AR(1) Residuals**	**Put AR(1) Residuals**	Model 4 Independent Variables *Summary Statistics for R(Call IV) and R(Put IV) when Model 4 defined event occurs in sample* **R(Call IV)**	**R(Put IV)**
Mean	0	0	0.2241	0.1954
Max	0.4778	0.4333	0.5546	0.5319
Min	-0.5712	-0.3595	0.0697	0.0665
Std. Dev.	0.1193	0.1038	0.1174	0.1106
Count	250	250	25	28

It is expected that the resulting estimated coefficients from the data defined by Model 4 have absolute values that are significantly larger than the corresponding values in the Simple Model. This fact would support the data specification for it increases the statistical relationship between returns on the underlying and implied volatility. It can then be assumed that the data specification succeeds in uncovering market dynamics in line with the thesis' overall assumptions. Also, it should be noted that the mean values of the independent variables in Model 4 are significantly larger at the 1% level than those in Model 3. This fact represents greater daily change in the variables for the tested days and thus may represent greater market microstructure dynamics that affect implied volatility. This then may lead to Model 4 having greater forecasting accuracy do its increased ability to uncover and analyze such affects in the options market.

Forecasting Models Regression Output and Analysis

Review of equations used in regression analysis:

1. Simple: $R(\text{Underlying})_t = \alpha + \beta_1\, R(\text{Call IV})_{t-1} + \beta_2\, R(\text{Put IV})_{t-1} + \varepsilon_t$
2. High IV: $R(\text{Underlying})_t = \alpha + \beta_1\, R(\text{Call H-IV})_{t-1} + \beta_2\, R(\text{Put H-IV})_{t-1} + \varepsilon_t$
3. Call vs. Put: $R(\text{Underlying})_t = \alpha + \beta_1\, R(\text{Call H-IV vs. Put})_{t-1} + \beta_2\, R(\text{Put H-IV vs. Call})_{t-1} + \varepsilon_t$
4. AR Shock: $R(\text{Underlying})_t = \alpha + \beta_1\, R(\text{AR Call H-IV})_{t-1} + \beta_2\, R(\text{AR Put H-IV})_{t-1} + \varepsilon_t$

In the tables on the following page, the estimated coefficients are shown for each of the independent variables with the percentage probability of each estimated coefficient equaling zero shown below. The goodness-of-fit value (R^2) is shown for each model in the far right column. The dependent variable in the models is the daily returns on the underlying (AMD or OEX). The independent variables that coincide with each forecasting model are shown on the top line.

AMD *Forecast Models*	Call IV_{t-1}	Put IV_{t-1}	C $H\text{-}IV_{t-1}$	P $H\text{-}IV_{t-1}$	$C > P_{t-1}$	$P > C_{t-1}$	C AR_{t-1}	P AR_{t-1}	R^2
Simple	0.031 24.2%	-0.033 23.9%							0.048
High IV			0.122 2.17%	-0.086 5.18%					0.059
Call vs. Put					0.097 2.21%	-0.119 2.13%			0.044
AR Shock							0.151 0.11%	-0.135 0.21%	0.064

OEX *Forecast Models*	Call IV_{t-1}	Put IV_{t-1}	C $H\text{-}IV_{t-1}$	P $H\text{-}IV_{t-1}$	$C > P_{t-1}$	$P > C_{t-1}$	C AR_{t-1}	P AR_{t-1}	R^2
Simple	0.028 0.95%	-0.015 4.56%							0.059
High IV			0.029 8.21%	0.007 53.8%					0.029
Call vs. Put					0.023 2.93%	-0.001 96.1%			0.033
AR Shock							0.024 15.5%	0.008 52.5%	0.022

For the case of AMD, the estimated coefficients' signs are all in agreement with the hypothesis. However, the Simple Model generates coefficients that are statistically insignificant from zero (probability of equaling zero is 24.2% and 23.9% for call and put implied volatility respectively). This result should not be surprising. The Simple Model incorporates daily data for the entire sample period. Since any success in the forecasting of stock returns is difficult at best, it is highly improbable if not impossible that success could be achieved for every trading day. This fact highlights the importance of correctly specifying changes in implied volatility in order to identify changes caused by market microstructure.

The significance and sign (positive for Call IV and negative for Put IV) on the coefficients in AMD Forecasting Models 2-4 suggests that certain changes in implied volatility today may contain information about the directional movement in the price of AMD stock tomorrow. This information and statistical relationship between the data will be the basis for refuting the Weak Form Market Efficiency Hypothesis.

The same results do not hold for the case of the OEX. All but one of the coefficients ($C > P_{t-1}$) are statistically insignificant from zero at the 5% confidence level. As discussed in Chapter 1, the OEX example is included as contrary evidence for the importance of a specific type of options market microstructure needed to affect implied volatility in such a way that daily changes in the data may and can reflect an introduction of information into the market.

Evaluation of AMD Forecast Models

Three different methods are presented in an effort to evaluate the Forecasting Models. Evaluation is based on the models' statistical accuracy and on their ability to provide a profitable trading strategy.

1. Statistical Analysis

The first method is a statistical analysis of forecast errors. The mean absolute error (MAE), the root mean squared error (RMSE), and the Theil Inequality Coefficient (TIC) are used to test the accuracy of the models. In order to evaluate the results, the statistical data is compared to the same tests performed on data form two time-series models of AMD returns. The two time-series are a one period auto-regressive, AR(1), and a one period auto-regressive moving average model, ARMA(1,1). If the market for AMD stock is efficient, then these two time-series models should be close to the best possible means by which AMD returns can be forecasted. The time-series models then provide a basis for which the thesis' models can be judged in relative terms.

AMD	MAE	RMSE	TIC
Forecasting Models			
Simple	0.03089	0.04359	0.85984
High IIV	0.03058	0.04311	0.80822
Call vs. Put	0.03123	0.04353	0.82841
AR Shocks	0.03045	0.04258	0.79596
Time-Series Models			
AR(1)	0.03021	0.44264	0.93572
ARMA(1,1)	0.03025	0.04425	0.94481

When different models are used to predict the same dependant variable, the one with the smaller MAE, RMSE or TIC is judged to be superior for forecasting purposes. From the table above, it can be seen that the implied volatility models provide more accurate predictions of daily returns on AMD than do the time-series models. This fact in of itself is not enough to refute the Weak Form Market Efficiency Hypothesis. However, it does provide evidence that there is a significant link between today's implied volatility and tomorrow's change in stock price and thus if nothing else it gives credibility to the Forecasting Models inherent assumptions and to this type of financial research.

2. Mean Returns

The second method used to evaluate the accuracy and validity of the forecasting models deals with actual returns on AMD and the hypothetical returns made possible by the models. The defined events in the each of the models can be equated to a *trade signal*. This signal can tell the trader when to buy or sell a stock in accordance with the thesis' hypothesis. For example, if the difference between daily change in call and put implied volatility is greater than one standard deviation above its mean, a buy signal will result and the trader can buy the stock on the close today with the expectation it will rise tomorrow. If there is a "shock" in put implied volatility and not in the calls (reasons for this are discussed later), a sell signal will result. This type of "real world" analysis of the models can provide the best evidence for or against their usefulness and thus whether or not they can be used as evidence against market efficiency.

The daily return on AMD following a buy or sell signal is compared to the mean daily return on AMD over the entire sample period. The daily mean return on AMD is considered here to be the *expected* daily return. Any statistically significant difference in the return from its mean value can thus be considered to be *unexpected.* The table on the following page reports the mean returns for the buy an sell signals for each of the forecasting models. The number of trading signals for each model along with the number of overlaps is also included. An overlap is defined as a day with both a buy and sell signal. The High-Volatility Model has an inherent weakness in that it may only identify events where actual volatility in the underlying is high. This weakness results in the particularly large number (50% of total signals) of overlaps. Simultaneous buy and sell signals are of obviously no use to a trader. For this reason, the mean return values for this model and for the AR Shock Model are calculated only for the days where an overlap does not occur. This process allows the test results to focus only on the days where demand/supply dynamics are the probable cause of changes in implied volatility and not actual stock price volatility.

AMD

Forecast Model	*Mean Return*	*# of Signals*	*Overlap*
High Volatility			
CALL	0.00351**	46	
PUT	-0.01211	45	
			23
Call vs. Put			
CALL	0.01148	25	
PUT	-0.02651	18	
			0
AR Shocks			
CALL	0.01836	18	
PUT	-0.02379	21	
			7

AMD SAMPLE MEAN RETURN = -0.00249

CALL = Buy Signal PUT = Sell Signal
** Mean Return insignificantly different from AMD mean return at the 5% level

The mean return following a specific event in implied volatility is in agreement with the thesis' hypothesis: model-specified events in call implied volatility forecast positive returns while put implied volatility can be used to forecast negative returns. Only the mean return for the Call High-Volatility model is insignificantly different from the AMD sample mean return.

3. Right or Wrong

The third method of evaluation looks at the number of correct signals. A correct buy signal produces returns the following day that are greater than AMD's mean return while a sell signal is defined as correct when the following day's returns are less than AMD's mean return. By using a simple binary calculation (correct = 1, incorrect = 0), a "successful" trading strategy should have a mean signal value greater than 50%.

AMD	# of Trades (Signals -Overlaps)	# Correct	% Correct
Forecast Model			
High Volatility			
CALL	23	10	43.48%
PUT	22	11	50.00%
Call vs. Put			
CALL	25	13	52.00%
PUT	18	14	77.78%
AR Shocks			
CALL	18	10	55.56%
PUT	21	18	85.71%

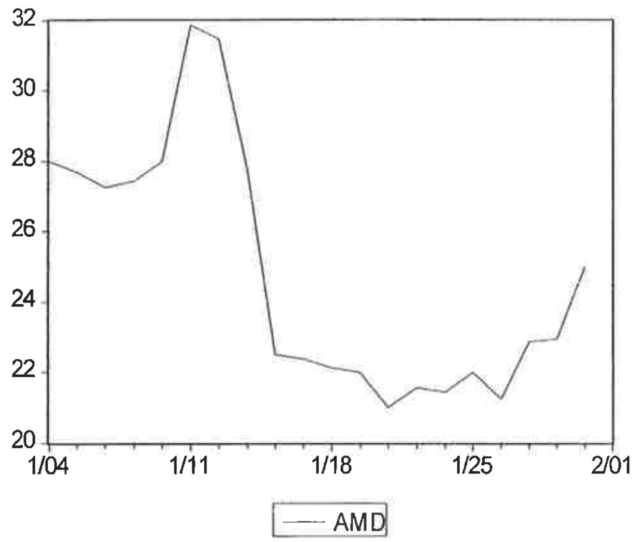
The above results, relative to each other, are in line with what was expected based on the results from the regression analysis. The High-Volatility Model is the weakest performer. 50% correct may be the best possible and logical expected outcome since results can only be associated with price dispersion and not directional movement in the underlying. Since the AR Shocks Model identified periods of changes in implied volatility that were significantly larger than those in the Call vs. Put Model, it was assumed earlier that this may lead to better forecasting results due to possible greater demand/supply dynamics in the options market. From the "% Correct" indicator, this appears to be the case in this evaluation.

The buy signals are less reliable than the sell signals and appear to be too close to the 50% barrier to allow a trader to be confident in their overall performance abilities. However, the sell signals appear not only to be accurate based on expected mean returns, but also on the overall amount of times they are simply correct and thus may be the best possible chance of refuting the Efficient Markets Hypothesis.

Chapter 4: Conclusion

A Final Story

AMD stock began 1999 rising in price in anticipation of a positive earnings report for the fourth quarter of 1998. From the graph below, it can be seen that when earnings were announced the morning of January 14, they were far below expectations. This new information entering the market caused AMD stock to suffer a one-day loss in value of almost twenty percent.



After the market assimilated this new information, the price of AMD stock traded in a relatively small price range. Instead of around thirty dollars per share, the market (post-earnings) valued AMD at around twenty-two dollars per share.

On Friday, January 29 a customer entered the options market and purchased five thousand February 22 ½ calls for an average price of 1.875 dollars. During the course of the trade (the customer placed five separate orders of one thousand contracts each), the price of AMD stock actually moved down a quarter of a point. The massive increase in demand for calls caused their price to increase even though the price of the underlying decreased. This irregular pricing dynamic resulted in the implied volatility of AMD calls exploding to the upside from fifty to almost seventy while the put implied volatility remained constant at fifty.

During a quiet, routine-filled day, a trade of five thousand, at-the-money, front month contracts results in numerous questions and intrigue amongst the market-makers. "Why is this guy buying so many calls?" "There is no news pending and AMD just came out with earnings." "I don't know but the calls sure were a good sale at two bucks with the stock below twenty-three."

On Monday February 1, 1999 the AMD options traders awoke to the following headline and story that answered their questions.

> "Gateway to use AMD Chips on New PC"
>
> NEW YORK, February 1 (Reuters) - PC maker Gateway Inc. will use Advanced Micro Devices Inc. AMD computer chips, instead of those of Intel Corp. INTC, on a new line of machines it plans to launch in March, PC Week reported. The North Sioux City, S.D. company, a long-time Intel loyalist, will use Advanced Micro Devices' K6-3 chip on the new line, sources told the weekly computer trade publication. Gateway plans to offer the new, AMD-equipped computer at a cheaper price than it could if it used Intel chips, PC Week said.

This definitely positive news for AMD resulted in the stock's price increasing over two dollars per share. The stock market's actions could hardly be deemed inefficient. On the close of trading Friday, the stock's price reflected all *publicly* known information and once this *new* information entered the stock market the price of AMD reacted accordingly. But if we were analyzing the options market using Forecasting Model 3 and 4, we would have noticed either a disparity in call and put implied volatility and/or a "shock" to the call implied volatility daily change time-series due to the sudden increase in demand. Using the tools provided by this thesis, the actual reason for AMD's future rise in price would not have be known, but a buy signal would have been generated that did in fact result in significantly positive, higher than expected earnings.

Concluding Comments

Finding inefficiency in a financial marketplace is a difficult task. If anyone were fortunate enough to find the perfect trading strategy, we would never hear about it. The lucky individual would be lounging on the beach, enjoying a cool drink with an umbrella sticking out the top and counting his millions of dollars. If inefficiency is found, it must be kept a secret. Assume for a moment that my models are one hundred percent accurate. As soon as the word gets out, everyone will be trading with them. But before you realize it, the opportunity to profit will be gone. The stock market will begin to react to the information in the options market immediately rather than some time in the future. When any risk-less, profitable trading opportunity is acted upon by numerous individuals, the market evolves and "learns" how to incorporate more information quicker resulting in a more precise pricing of the underlying. Maybe someone will make a quick dollar, but sustained risk-less profits are essentially impossible.

Take for example the case of Long-Term Capital. Long-Term Capital is a hedge fund whose leaders (two Nobel Prize winners in economics) believed they developed a forecasting model that generated guaranteed profits. In the beginning, their model resulted in extremely high returns. But in the late 1998, market conditions changed dramatically and the fund lost nearly four billion dollars (most of which was highly leveraged by banks and brokerage houses). Even the most gifted traders and economists can fail miserably in attempts to find a foolproof trading strategy that provides higher than expected profits over the "long-term".

It was foolish and possibly arrogant of Long-Term Capital to rely so heavily on a purely quantitative market analysis for its decisions. I would be even a bigger fool if I did the same with my models. But from an academic standpoint they do appear to provide evidence of a statistical link between information contained in today's options prices with tomorrow's price on the underlying. The assumptions and the logic are quite straightforward. Options by nature are forward looking. This combined with the fact that equity options have become an increasingly popular investment tool leads to the hypothesis that it is possible for options markets to contain information that the stock market does not. The trick is to find a model that identifies and deciphers this subtle information in the price of options.

The statistical significance and the signs on the estimated coefficients in the AMD Forecasting Models regression analysis provide the best evidence that call and put implied volatility related to expected future price dispersion *and* to expected future directional movements in the underlying. The OEX data is presented to illustrate that even if all of the assumptions hold for AMD, they do not hold for all types of options markets. The differentiating factor between the two markets is their microstructure. The market must be such that the forces of supply and demand can significantly affect the options pricing structure in ways that are independent of other Black-Scholes options' price determinants.

In the mean return analysis of the Forecasting Models it did appear that they might have the ability to provide consistent, higher than expected returns over the long run. A buy signal resulted in returns higher than the sample mean AMD return and the sell signal resulted in lower returns. The percentage of correct signals leads to the conclusion that while profitable, the models remain a risky strategy. If the percentage correct is only slightly higher than fifty percent, one extremely bad signal will wipe out all previous gains. But how many times out of a hundred do you have to be right in order to deem a trading strategy useful and thus refute the Market Efficiency Hypothesis? In the case of Long-Term Capital, they might have been right the first ninety-nine trades and then on the hundredth trade they lost everything and more.

If it is difficult to find a risk-less trading strategy, it may be even more difficult to actually conclusively prove that a risk-less, profitable strategy actually refutes the Efficient Markets Hypothesis. As soon as someone thinks they have all the answers, the market changes and new theories must be formed. However, the models in this thesis do provide some evidence against the Weak Form Market Efficiency Hypothesis. The weak form states that past pricing data information cannot be used to forecast returns. The assumption for the most part has been that "all past information" refers to information in the stock market itself. Derivatives (options) are "derived" from the stock market. It follows that data from this market should also be included in any test of stock market efficiency.