

Comparing Historical and Implied Volatility Estimates in Efficient Portfolios

Isaac J. Faber¹

Abstract

This paper evaluates the performance of efficient portfolios with differing sources of volatility estimation. One of the primary assumptions of modern portfolio theory is that the parameters, asset means, standard deviations and covariance, are known. In practice these values are not known and have to be derived from reliable and accurate sources. Two sources of volatility estimation are compared in this paper; the classical statistical approach and Black-Sholes implied volatility. The time horizon preceding, during and after the financial crisis of 2008-9 is used to evaluate the two sources of information by generating competing monthly tangency portfolios. The results suggest that while historical statistical estimates outperform there is no statistically significant difference between the two over the given time horizon. However, this is partially due to there being a tendency for implied volatility to remain over inflated beyond the crisis into the recovery period. This then leads to significant losses in portfolios that solely use the implied volatility measure.

JEL classification numbers: C61, G11, G12.

Keywords: Implied Volatility, Efficient Portfolio

1 Introduction

In today's financial landscape there is a diverse set of securities which an investor can choose from given their preference for risk. One of the major dilemmas for an investor is how to match their risk tolerance to that of the security or securities in which they choose to invest. It could be argued that an investor can have a great understanding of the amount of risk they are willing to take on; however, selecting an investment which performs to that risk threshold is much more difficult. Portfolios are constructed given an understanding that risk and return are strongly positively correlated. This means that the

¹Systems Engineering, United States Military Academy.

greater return an investor pursues the greater risk they must accept. Part of the trouble an investor encounters is how to know how much risk they are actually taking on when they purchase an investment. This is a problem because securities may not behave in the future as they have in the past. In many disciplines, such as physical sciences, measured historical statistical performance is an excellent indicator of future performance. However in social sciences, such as finance, making this assumption about historical performance may not be the best approach to estimating risk for an investor.

A potential exists that historical estimates can lead to too much or too little risk for a particular investor. One of the major assumptions in modern portfolio theory is that financial performance can be quantified in terms of the mean and variance of a given security. This implies that in order to understand the return and risk trade-off of an investment is an investor simply needs to know the mean (return) and standard deviation (risk) of returns as well as the covariance between portfolio components. The contribution of this work is to evaluate how well these types of portfolios perform with only differing volatility estimates. The characteristics of an efficient portfolio in which an investor can achieve the greatest return for the least risk under these conditions is:

$$\mu_p = \mathbf{m}\mathbf{w}^T \quad (1)$$

$$\sigma_p^2 = \mathbf{w}\mathbf{C}\mathbf{w}^T \quad (2)$$

The mean and variance are known as a function of portfolio asset weights (a vector \mathbf{w}) which are the portion of an investors total wealth, means, variance and covariance (represented as a matrix \mathbf{C}).

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \sigma_{2,1} & \dots \\ \sigma_{1,2} & \sigma_2^2 & \dots \\ \dots & \dots & \dots \end{pmatrix} \quad (3)$$

The weight vector ($n \times 1$) is found using the following program where the desired return is used as a constraint and the sum of weights is one:

$$\text{Min} : \sigma_p^2 = \mathbf{w}\mathbf{C}\mathbf{w}^T \quad (4)$$

$$\text{s.t.} : \mathbf{w}\mathbf{1} = 1 \quad (5)$$

$$\mu_p = r \quad (6)$$

On the face this seems fairly straight forward, but given the dilemma just presented about historical information, estimating these parameters may not be as reliable as preferable. Typically major academic texts on the topic of portfolio theory use historical statistical estimation (referred to as the Classical approach) to construct portfolios applying approaches developed by Markowitz [1]. Parameters are estimated from a sample of n historical observations using the following:

$$\hat{\mu}_i = \frac{\sum_{i=1}^n r_i}{n} \quad (7)$$

$$\sigma_i^2 = \frac{\sum_{i=1}^n (r_i - \hat{\mu}_i)^2}{n-1} \quad (8)$$

$$\sigma_{i,j} = \frac{\sum_{i=1}^n (r_i - \hat{\mu}_i)(r_j - \hat{\mu}_j)}{n-1} \quad (9)$$

This is done typically to demonstrate the techniques of how to construct portfolios and not necessarily how to select actual portfolios for investment. So the question becomes, what is the best way for an investor to go about understanding the actual risk they are taking on when purchasing a security? The purpose of this paper is to explore an alternate technique for looking at options to see whether it is useful in understanding the risks in which an investor is taking. When an investor is interested in purchasing a security, and more specifically a common stock, there are often related securities which can provide information about risk specific to this investment. By looking at the prices of other related securities that derive their value from the stock, in this case options, an investor has another source of information from which to judge risk besides historical performance. Stock options are derivatives of a stock because their value is derived from the stock. These derivatives are valuable because they are contracts based on market prices that will be realized at some point in the future. This characteristic makes options forward looking, and therefore, a representation of other investor's sentiments about the future of the stock. This is important because in an efficient market, all known information is reflected in the prices of securities, including options. This feature makes options prices a potentially valuable source of information about risk. Basically, if an investor wants to understand the risk associated with a particular stock they can look at the behavior of other investors to determine the overall market sentiment about the risk that they (the investor) are concerned about. However, the question of whether this source of information is in fact better than the historical estimation discussed earlier needs to be addressed. It is this comparison that constitutes the contribution of this paper. The approach described in the following sections compares the similarities and differences between historical information and derivative based information in dealing with portfolios. In particular testing which has a better track record in terms of performance. This paper will use as its test time frame the financial crisis of 2008. A series of efficient portfolios will be generated using historical and options based risk information and then empirically compared. Any risk measure used during this time frame can be evaluated it in its ability to predict and them perform through a significant market event.

2 Background

In recent years, the financial derivatives industry has experienced an enormous boom. Derivative security trading today (including futures, forwards, swaps, and options) dominates all other securities in size and scope. Derivatives by definition gain their value through an underlying asset such as a stock (as stated earlier), bond or commodity. As such, the value of most derivatives which have future exercise dates can be used to,

generally, derive a future market worth of such an asset. In the field of stock options the prevailing approach for understanding and evaluating prices is the Black Scholes formula (BS). This formula was published in 1973 by Black and Scholes [2] and serves as the seminal options pricing model. The model is designed to give a valuation of ‘European’ style call options ($C(S,t)$, meaning they can only be exercised upon expiration) as a function of five parameters within a set of assumptions. The parameters are the current underlying asset price ‘ S ’, the strike price ‘ K ’, the time to maturity ‘ $T-t$ ’, the riskless rate of return ‘ r ’, and an instantaneous volatility parameter (which describes the instantaneous standard deviation of the returns) ‘ σ ’. A quick glance at the models parameters reveals that the first four are known with relative, or complete, certainty and the fifth, volatility, is unknown. This parameter is the same unit as the statistical estimate from equation 8. The formula takes the following form:

$$C(S,t) = N(d_1)S - N(d_2)Ke^{-r(T-t)} \quad (10)$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \quad (11)$$

$$d_{21} = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \quad (12)$$

Another important consideration is that BS cannot produce valuations for put options ($P(S,t)$), these prices are derived from the put-call parity.

$$P(S,t) = Ke^{-r(T-t)} - S + C(S,t) \quad (13)$$

$$= (1 - N(d_2))Ke^{-r(T-t)} - (1 - N(d_1))S \quad (14)$$

Using historical estimation of volatility in the BS formula, results in a skew in valuation from actual market prices verses the model’s predicted values. This price skew is often referred to as the ‘volatility smile’ and is the source of a major field of study in the financial derivatives literature. Not long after the valuation differences were observed did the financial community begin using BS as a method to ‘back out’ the volatility. This backing out (rearranging BS and solving for volatility using the market price as the value) results in a value known as implied volatility (IV). This value will form a significant portion of the basis of this paper and will be explored as a potential alternative measure of risk. It is important to highlight that this value as used here) is model dependent and therefore inherently biased. To communicate this dependence, often times IV is referred to as ‘Black-Sholes implied volatility’ [3]. Implied volatility as a predicative parameter will be explored as a major component of this paper. As addressed in the introduction, once the volatility skew became recognized from the BS model, practitioners and researchers began to back out the implied volatility in both an attempt to explain it and an attempt to exploit it. The major problem is that the parameter cannot be calculated explicitly (in closed form) and must be estimated. For example, one of the methods for backing out the IV was published by Manaster and Koehler [4] where they employ existing techniques (developed by Newton) to approximate the exact value. This approach is a type of bracketing procedure. As these techniques became more wide spread IV

became more accepted. In today's financial landscape IV is a common piece of information typically accompanying an options price quote.

As the options literature progressed, there became a distinct split where many publications sought to adjust the options formula to account for the volatility and others sought to explain or test the value of the parameter. This is basically the difference in those who think that the BS model can be modified to represent reality and those who think that the model can be used to derive information about reality. In the literature this BS implied volatility, as measured over all strikes for a given expiration, has been referred to as the 'volatility smile.' The reason for this is that way out of the money strikes trade at much higher prices than is predicted by BS creating a sort of smirk when viewed on a graph. These prices are skewed negatively in that on average equivalently out of the money puts are more expensive than calls, and have been significantly so since the 1987 crash [5].

The recognition of the smile led to more publications concerning an entire implied volatility surface or IVS. This surface is a representation of all implied volatilities at every available strike and all expirations as well. Often functions are fit onto this surface in order to assist in the pricing of derivatives, which do not have robust options chains. As is obvious this topic of implied volatility is quite active in the literature. The reason for this is simple as stated in Fengler [6] "unlike volatility measures based on historical data, it [IV] should reflect market expectations on volatility over the remaining life of the option." Basically IV provides forward looking information, even if biased, about the underlying security. Contrast this with historical estimates which only included information about the past. There are, of course, many ways to estimate the future volatility of a security and IV is just one of them. As mentioned in the introduction, historical statistical estimation (calculating a value given historical data) is often taught. However, there is another important method of volatility estimation, which has a large presence in the literature as well, called the ARCH models. ARCH models (autoregressive conditional heteroskedasticity) exploit and represent the realization that volatility tends to 'cluster.' This basically means that some fairly reliable models can be built to predict future variation based on historical data [7]. However, research done by Poon and Granger [8] compared various 'volatility forecasting methods' and found that IV tended to be superior and that ARCH models tended to even under perform historical statistical estimation. Research comparing IV with realized volatility has also been active for quite a while.

The research that has been done in the use of implied volatility as a predictor of realized volatility began shortly after the publication of the BS model, with the earliest work being done by Lantane and Rendelman [9]. In this paper the authors calculate the implied volatility and then take a weighted average over various strike prices to estimate future volatility. This work was followed by a number of others, each with their own interpretation of the uses of IV. For example Chrias and Manaster [10], Schmalansee and Trippi [11] and Shu and Zhang [12], all confirm the validity of the use of IV as a good predictor of future volatility. There are other authors such as Fleming who demonstrate that IV is upward biased but still a good predictor [13]. Also Blair, Poon, and Taylor [14] show the value of the use of IV in terms of its incremental informational content. The general consensus of the literature is that implied volatility is a good, if not the best at times, indicator of realized volatility (an exception being Canina and Figlewski [15]) as demonstrated by Poon and Granger. Another interesting and related recent paper by Zing, Zhang, and Zhao demonstrate the predictive ability of the smile as it concerns asset

returns [16]. Many of these topics are also well summarized by Fengler in his book *Semi parametric Modeling of Implied Volatility* [6].

As is demonstrated by the literature, IV seems to be a superior source of information about the future volatility of an underlying asset. But the question remains as to how much of a benefit it actually is, or even how to arrive at a singular value, as there are varying IV values for each strike at each expiration. For the concern as to how beneficial the estimate is, each of the cited publications gives comparisons to predictions and realizations but this may not be the specific information that an investor wants. Also, because IV is calculated for a specific option (a fixed strike price at a particular expiration) how or what is the investor to use as their parameter for estimation? For example Lantane and Rendelman [9], use various weights to value each IV and then calculate a single value for a given expiration by averaging all of the available IVs across the strikes. There are many methods for estimating the IV parameter. Some of the literature simply uses the 'at the money' option IV and ignores the rest, others incorporate varying strikes but not all. A critical point of the methodology of this paper will be to decide which parameter to use. Because IV needs to be averaged or estimated, this could have a potentially large impact on the outcome. Using IV to develop a parameter is the next area that will be addressed and will lead to an investigation of the volatility index or VIX.

The VIX is the short term volatility index tracked by the Chicago Board Options Exchange or CBOE. This index is often referred to as the 'fear gauge' [17] because it measures the implied volatility of the S&P 500. As conventional wisdom goes; the greater the market fear the higher the index will go. Analytically this can be interpreted as the greater the perceived volatility (prices of options) the higher the index goes.

The index is a weighted average of short term (defined as 30 days) implied volatility of actively traded options on the S&P 500. Of course, there is not always an option chain that expires in thirty days so the index is most often a combination of two weighted averages derived from the options chains that expires in less than thirty days and the options chain that expires between sixty and thirty days. The VIX 'value' as reported by the CBOE is given in units of annual percentage of continuously compounded returns, and is representative of standard deviation from BS (though the VIX is now calculated using a model independent version) [18].

The VIX index has become a staple for both the use of practitioners and in financial literature. The VIX was introduced in 1993 by Robert Whaley and has become the standard for measuring market risk sentiment about the future of the market in general [19]. When the index was first introduced it was based on the S&P 100 and has since been revised in 2003 to represent the S&P 500 as well as 'transformed VIX from an abstract concept into a practical standard for trading and hedging volatility.' according to the CBOE.

There are also several other similar indexes tracked by the COBE, such as the VXN which tracks short term implied volatility of the NASDAQ exchange and the VXV which tracks 3-month implied volatility on the S&P 500. However, for practical and academic purposes the VIX gets the most attention. This is likely because the S&P 500 is considered the best available market proxy (not the best quality but the best in terms of convenience and quality) and short term volatility is often of most interest. The popularity of the VIX led to the creation of many new financial instruments based on implied volatility including VIX futures, beginning in 2004 and VIX options beginning in 2006. However, it is important to point out the difference of these instruments and the potential uses of the information that the VIX provides.

Trading on the ‘value’ of the VIX is a venture into an investment in the markets perceived sentiment of future volatility and not actually trading in volatility. So if a future, a type of derivative investment, on the VIX is purchased, an investor is making a choice about what they think the market sentiment will be in the future not necessarily what the market itself will do. This type of decision is outside the scope of this paper so trading in VIX based instruments or even implied volatility will not be further explored. However, the question as to the success of the VIX (and by extension implied volatility) in predicting variation in the S&P 500 is important to this research. Figure 1 shows the daily values of the S&P 500 against the monthly values of the VIX over the previous decade.

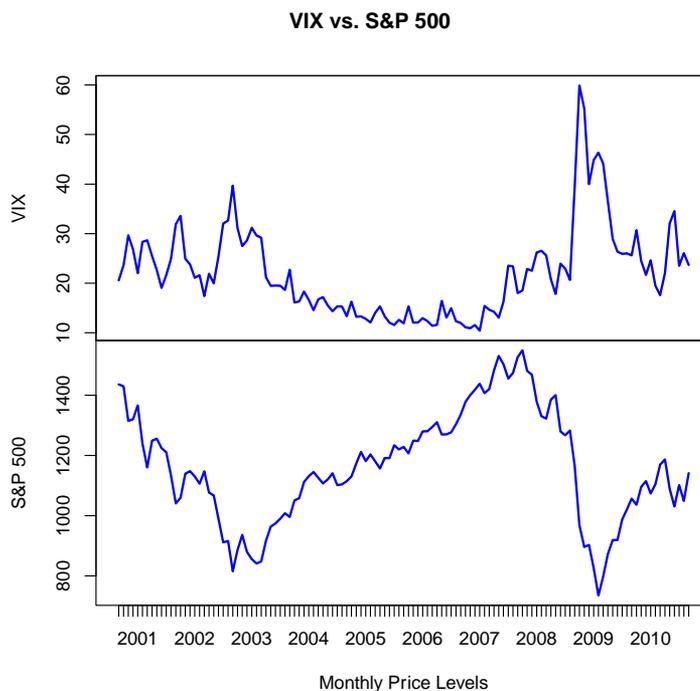


Figure 1: Price Levels of VIX vs. S&P 500

Just from a quick glance we can see that there does seem to be correlation between several of the VIX dramatic increases and the S&P 500 declines (and vice a versa) such as the recent market crash in 2008-9. The major question of interest is whether or not this casual observation is accurate. The literature on the predictive abilities of the VIX (and its counter parts) is significant and began shortly after it (the VIX) was introduced and continues to be an active area of research. For example Corrado and Miller (2005) evaluated various VIX related indexes predictive power and found them to have significant forecasting ability as well as improved quality since 1995 (early versions of the index had complications with various errors). Also Banerjee, Doran and Peterson [20] tested VIX related variables concerning portfolio returns and show that they have strong predictive abilities.

The popularity of the VIX and its predictive abilities has even extended to more general work such as Scheffer et al. [21] who study general predictors of critical transitions in complex systems using the index as a case study. However, the VIX is not perfect and a

paper by Simon [22] shows that the VXN has strong predictive abilities but trades at much higher levels than realized volatility (7.5% in this study). This result can be interpreted as the VXN and by extension implied volatility may be better explained as investor sentiment and not as a strict predictor of future variation.

This, of course, makes sense intuitively as an investor purchasing an option is not explicitly making a trade on volatility but instead based on either hedging preferences or sentiment. This notion is further confirmed by Kaplanski and Levy [23] who show that implied volatility in the market increases after significant disasters, specifically aviation related disasters, without any corresponding increase in realized volatility. So a reasonable interpretation of the literature on implied volatility is that it is a good predictor of future volatility but does in fact represent investor's sentiment and not direct bets on future volatility. In this light the description of 'the fear index' may be the most accurate. As we can see from the literature implied volatility has some significant benefits and enjoys an excellent reputation as a predictive tool. The major topic that this paper will test is how it (IV) performs in terms of selecting a portfolio. This approach will yield not only a test about the other findings in the literature, but will also give an indication of the quality of implied volatility as a parameter in terms of realizing returns.

3 Data

The selection of appropriate data for the selection of well diversified portfolios is important. Most of the research reviewed in the previous few sections had their focus on one particular index, the most common being the S&P 500. The difference in this paper is that the focus is being placed on a portfolio. This, of course, requires multiple securities to choose from in order to create a set of appropriate weights. This is a challenge in that there is a danger of choosing too narrow a selection of securities, for example only choosing five individual stocks. In general it is preferable to select a set of securities that will best represent a broad sentiment of market conditions. In this light the data for this paper will be represented by five individual exchange traded funds or ETFs. The purpose of an ETF is to match the performance of something that is not readily tradable on the stock exchanges. For example one of the ETFs that I will be using tracks the S&P 500 and trades on the New York Stock Exchange (NYSE) under the ticker symbol SPY. This is of course important because with this one fund I can represent five hundred companies that constitute the index. I will be making assumptions about the ETFs and their liquidity in that they will be a direct representation of the securities which they claim to reflect. So any timing, tracking, or fee related errors will be ignored. This assumption is supported by the literature for index based ETFs and does not apply to more exotic ETFs (such as leveraged or futures based funds) which are beyond the scope of this paper.

In addition to the diversity and quality of ETFs available for tracking funds, the other critical feature for this research is the availability of robust options chains. This is important in that the quality of the parameters that will be used to test implied volatility will need to be derived from the publicly traded options. For example SPY the previously mentioned ETF that tracks the S&P 500 has, at any given time, up to seven available options chains with expirations out to several years. If a fund has a small amount of available options chains then it would be a poor candidate and would not lend itself to this particular test. It is only relatively recently that ETFs have gained investor attention and by extension options availability and liquidity. Using this new information is another

important contribution of this paper. The data itself is collected from DeltaNeutral a reputable vendor of options and stock data. DeltaNeutral has provided support for not only academic research but to banks, and hedge funds as well. The source of the information is important because its reliability is paramount to the analysis of the thesis. In order to double check the stock data provided by DeltaNeutral a limited cross reference with available data from CRSP was used. CRSP or the Center for Research in Security Prices is the premier database for stock data and is widely accepted as the most accurate and comprehensive set of data for this regard. The only reason that CRSP data is not used exclusively in this regard is that the author's access does not include options information. Every set of data has to address several complications of prices as they mature. Over time, the two most common of a stock's price are splits and dividends. Because a stock quote is a stated price for only one share of a company's stock if a company does something that directly affects the price or value of the share at any point in time then that event must be reflected in the stock price looking backward. For example if a company issues an annual dividend of five percent per share then anyone owning the share would have five percent greater value regardless of stock price. So looking back from the point of this dividend issue, the price of the stock should be discounted accordingly. The data used for this paper has been adjusted for such events and is reliable for testing and research purposes. In keeping with the principals of diversification and available options chain information the following five ETFs have been specifically selected: SPY, QQQ, AGG, IJR and VNQ. The funds represent the large cap stocks, NASDAQ, bond markets, small cap stocks and real estate respectively.

The data that will be used spans a range of six years from January 1st 2005 to December 31st 2010. This data set covers several significant market events including a bull run through 2007 and the subsequent crash in 2008-09 as well as other notable occurrences such as two presidential elections. This range will offer an excellent test for any estimate of future returns. It is useful to look at the performance of the components of the portfolio over this time period in this illuminates some of their defining characteristics. In order to evaluate each of the components, the distribution of monthly returns during this period as well as the pair wise comparisons between the securities will be addressed. These factors are important, of course, because they are the parameters that will be employed to construct the portfolios using the Markowitz algorithm. Figure 2 shows the individual components prices movements over the given range.

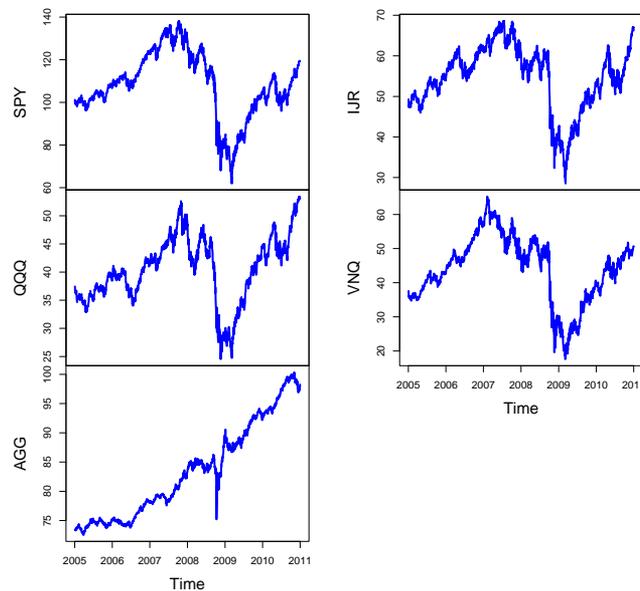


Figure 2: Component Prices Over Tested Horizon

We can see from these price changes that all of the components have had several significant events most notably in 2009. The monthly returns of AGG reacted differently to the market crash than did the stock based ETFs. In addition, it is also useful to look at the time series of monthly returns for each of the components as shown in Figure 3.

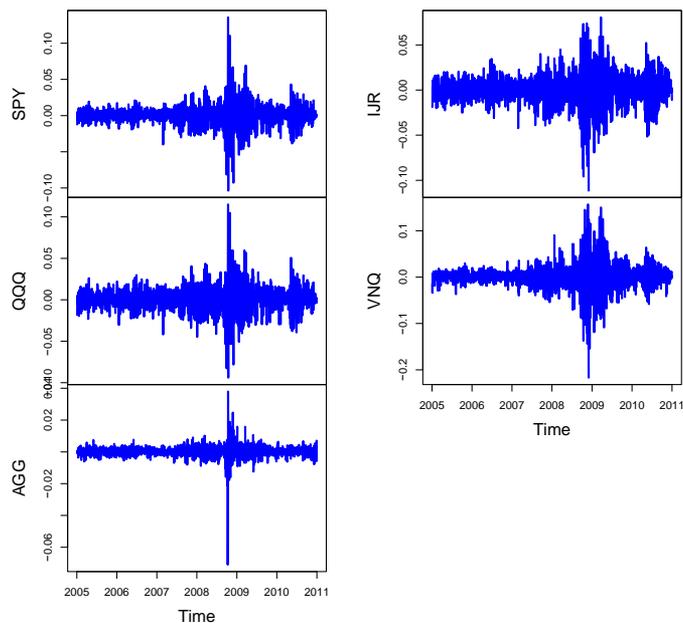


Figure 3: Component Returns Over Tested Horizon

In a similar fashion we can see a significant range of volatility in the returns in each of the components over the given range to be evaluated. The notable exception to the negative volatility reaction of returns is the bond tracking ETF of AGG which shows a significant positive event during the crisis. One of the questions that will be addressed in later sections is whether implied volatility can be leveraged to anticipate such moves and, if this is the case, we should see an increase in the weighting of bonds before this event. In addition to the individual securities performance the pair-wise comparisons are represented by the scatter plot in Figure 4. This plot shows that the stock funds seem to have a positive relationship. This is to be expected as the funds contain several of the same component companies. However the bond fund has a less significant relationship to any of the other funds, which makes for more robust diversification. In addition to the graphical representation of the data the Table I depicts the four empirical moments of the historical distribution parameters as they are calculated explicitly. This table demonstrates some important considerations. First is that none of the means has a negative value (though these values are not stated with any confidence).

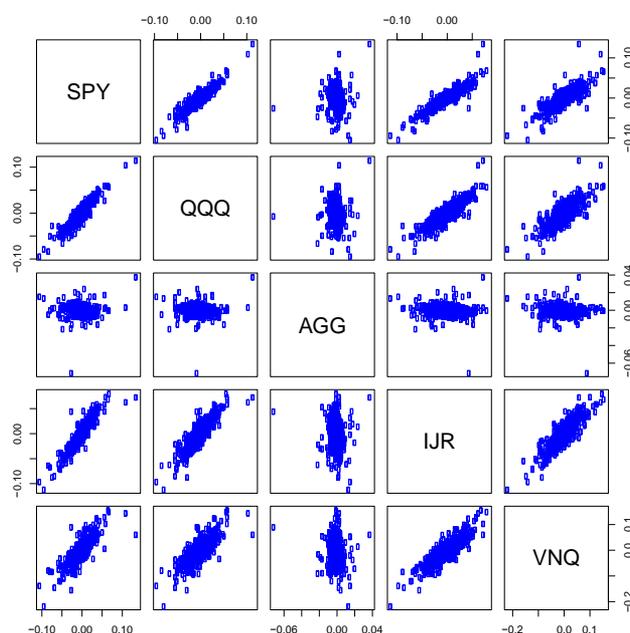


Figure 4: Pairwise Comparison of Component Returns

In the methodology section we will discuss the importance of the mean values in determining the weights using the Markowitz algorithm (equations 1 and 2) but for here it is important to note that any negative mean values could bias the portfolio against that particular fund over the entire tested period. Second the fund that represents the NASDAQ has the best mean value at 0.6% monthly which could have a great influence on its weighting.

Next we can see that the standard deviations are fairly consistent across the funds save for the bond fund which has a value over half that of the next closest fund (at 1.27%). Also the skewness and excess kurtosis values (calculated as coefficients) are worth testing to see if the distributions are significantly normal (a key assumption in much of the portfolio

literature). This final column in Table 1 show the result of the normality hypothesis test at a significance of .95 using a JB test.

Table 1: Historical Empirical Moments of Data

	Mean	Std Dev	Skewness	Kurtosis	JB Stat
SPY	0.002434	0.048839	-1.00155	1.750602	21.23099
QQQ	0.005658	0.06129	-0.70257	0.466179	6.575235
AGG	0.004038	0.012738	1.211291	5.519332	108.9958
IJR	0.004246	0.062164	-0.74652	1.492749	13.3724
VNQ	0.016754	0.052638	-0.54452	1.284529	8.508093

As we can see, we reject the null hypothesis of normality for any of the funds in this data set. However, for portfolio construction purposes we will proceed with the assumption of normality. In addition to the individual distributions data, table II shows the covariance between the funds: We can see from the table the relatively high covariance between the stock funds and the low/near zero values between the bond and stock funds. This is confirmation of the observations from the pair-wise comparisons in Figure 9. There is a near zero relationship between bonds and the S&P 500. These low values are valuable for the sake of diversification in efficient portfolio construction. In addition to the performance of the funds themselves, the other and much larger component of the data is the options chains associated with them. The options data for each fund spans four years, which allows for a testing (warm-up) period of 24 months in developing a statistical baseline, and covers each of the strike prices associated with the fund at the time of the monthly quote.

Table 2: Covariance Matrix of Returns

	SPY	QQQ	AGG	IJR	GLD
SPY	0.002385	0.002694	0.000096	0.002802	0.000223
QQQ	0.002694	0.003756	0.000029	0.003299	-9.9E-05
AGG	0.000096	0.000029	0.000162	0.000065	0.000215
IJR	0.002802	0.003299	0.000065	0.003864	0.00007
VNQ	0.000223	-9.9E-05	0.000215	0.00007	0.002771

Because this data set records monthly returns relative to the closing price on the first trading day of each calendar month, the options data must be recorded from that point as well. This collection nuance creates challenges with interpreting the options data which will be addressed in the methodology section. The amount and nature of the options data makes any summarization of it difficult. In fact, an attempt to do so is unnecessary for the scope of this paper. However, a general assessment of the data is reflected in Figure 1 where this VIX index is compared to the S&P 500. The VIX is, in fact, a representation of short term implied volatility (30 days) and has a value comparable to the value associated with the SPY fund. The other stock funds will likely follow this general shape and characteristics of the VIX. However the bond fund may not have as direct a comparison, which we would suspect as the correlation between them and the stock funds is lower, as observed in the pair-wise comparisons.

In summary the data set covers six years of information on five funds which represent a wide variety of equity, debt, and real estate. This breadth of selection should result in a well diversified portfolio which should be a good approximation of the market portfolio. This is important as a portfolio with little diversification is potentially susceptible to bias and would not be as useful of a test for this papers stated purpose.

4 Methodology

In the background section the predictive abilities of implied volatility where reviewed. The general conclusion is that it is likely the best available metric for predicting future volatility. While there are several publications that dispute this claim they are in the minority. The other papers referenced so far have shown that while IV is a good predictor it also lends itself to overreactions and is generally overstated. The contribution of this paper to the literature shows implied volatility in the context of constructing a portfolio. The prime difference is that comparing statistical estimation and IV against realized volatility does not give an indication of how much better either metric is in terms of predicting portfolios performance.

When evaluating various portfolios, where the only difference is origin of the volatility parameter, we can uncover how well either does in realizing gains or losses. We can look at how well each of the selected portfolios performs over time by comparing their returns directly. We can also study the weights within the portfolios. For example, if a crash is anticipated in one area, the weights in the portfolio should shift away from that security.

The first step in the methodology in testing various portfolios is to develop the implied volatility parameter. With historical statistical estimation, a single parameter is the product of analysis of the historical data (the squared deviations). In this paper we use a rolling 24 month horizon of log adjusted continuously compounded returns. However, with implied volatility it is not as straight forward. As discussed in the data sections, we are taking the prices of the assets from the options expiration data in order to have neat representation of the holding period of one month for which we are looking. If data was selected from another day then the volatility from two options chains would have to be balanced in order to adjust for the time horizon correct. This is what the VIX calculation does for example. The VIX represents the short term implied volatility or thirty day where this paper seeks for a monthly holding period (regardless of days for that particular month).

In order to compensate for this extrapolation nuance, our data is collected only when options have expirations that can be measured in months. This is done by calculating the prices of the securities on the first trading day after that months option has expired. For example, if the monthly option expires in June of 2008 on the nineteenth (the third Friday of every month) then we take the values from the twenty-second or the next trading day. All options values then only have to be calculated for the next expiration which is one trading month away. The other major issue is that there is a range of strike prices for all available expirations. These ranges all have a unique IV associated with them for both puts and calls. Some research simply uses the at the money option which is the approach that I will use. The average of the two options implied volatility, the put and call, at the money follows this form:

$$\sigma_{IV} = \frac{\sigma_{Put-IV} + \sigma_{Call-IV}}{2} \quad (15)$$

The actual calculation of implied volatility is done with Newton's method where the price is taken from the options data at the midpoint of the bid-ask spread and the risk free rate is taken from that days three month Treasury bond yield to maturity. This takes the form of:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (16)$$

Where numerator is the BS formula solved for the options (equation 9 and 10) price minus the market price equal to zero and the denominator takes the form:

$$f'(x_k) = \frac{1}{\sqrt{2\pi}} S e^{-qT} \sqrt{T} \exp\left(-\frac{(d_1(x))^2}{2}\right) \quad (17)$$

This recursion is started with a value of twenty-five percent and repeated until two successive values are within 10^{-6} .

There are numerous methods for selecting portfolios; however, for the purpose of this paper we use the Markowitz algorithm discussed in the introduction. This approach gives an intuitive and well understood selection of securities based on their expected returns, volatility, and covariance. The algorithm has been well studied and has several promising attributes. The most important facet is that it has been used to prove that if two or more securities which have anything less than perfect positive correlation then improvements to risk exposure can be gained by diversifying. The algorithm takes a set of candidate securities and their individual and pairwise values and generates optimal weights for risk exposure and return. The various combinations of weights possible between each security available are the different portfolios.

Each possible portfolio has an expected return and variance associated with it. This combination of characteristics is typically expressed as a trade of risk and return and is often referred to as the 'bullet graph'. The benefit of the Markowitz algorithm is that it lets us identify a portion of this graph as the efficient frontier. The algorithm is used to derive certain combinations of weights based on the inputs described above to generate this frontier. The frontier is a set of portfolios that dominates all others in terms of risk (volatility) vs. reward (expected return). This is a case of Pareto optimality where every possible portfolio along the efficient frontier is 'better' in terms of return for lower risk than any other possibility. The efficient frontier is a convex combination of any two points on the frontier, so constructing it requires identifying any two of the possible frontier portfolios. This is done through a simple optimization problem. The first step is to identify the global minimum variance portfolio, which is the portfolio which gives us a set of weights that has the lowest possible variance given the set of securities from which to choose (done by applying expression 4 and 5). The second step is to find another portfolio with a minimum variance given a specified return (done by applying the constraint in from Figure 6). Next, these two portfolios can be used to construct the entire frontier. It has been shown by Markowitz that if there is no restriction on the weights in the portfolio then the frontier is a convex combination of the mean and variance and any

portfolio can be found with the following equation:

$$\mathbf{z} = \alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \quad (18)$$

Where \mathbf{z} is the weight vector of any efficient portfolio and \mathbf{x} is the global minimum and \mathbf{y} is minimum variance with the constrained expected return. However, for the purposes of this paper, there is only one special case portfolio which we are concerned about, the tangency portfolio. The Markowitz algorithm was expanded by Sharpe to include the potential of a risk free investment, often considered to be the rate of return on a bond issued by the U.S. Treasury (for this paper I use the three month T-Bill). Because the bond is considered risk free its variance is assumed to be zero. Sharpe showed that with a balance of a risk free asset and the tangency portfolio, any portfolio on the frontier can be dominated (achieves the same expected return with lower variance or visa versa).

The tangency portfolio is the portfolio that is on the point where a tangent line intersects the efficient frontier with a straight line drawn from the risk-free the point. Sharpe's work has also been expanded into the capital asset pricing model, or CAPM, which holds that if all available investments are considered, the market portfolio is the tangency portfolio and that all investors will hold some combination of the risk free asset and the market. The line which demonstrates this relationship is called the security market line or SML. The tangency portfolio is calculated in a similar manner to the GMVP, which is by applying an optimization problem. However, instead of minimizing the variance, the tangency portfolio seeks to maximize the 'Sharpe's slope' of the frontier. Sharpe's slope is a measure of risk that incorporates the risk premium return and the standard deviation. The expression is:

$$\text{Maximize: } \frac{\mathbf{t}'\boldsymbol{\mu} - r_f}{\sqrt{(\mathbf{t}'\mathbf{C}\mathbf{t})}} \quad (19)$$

$$\text{s.t. } \mathbf{t}'\mathbf{1} = 1 \quad (20)$$

Like the frontier, any portfolio containing a combination of the tangency portfolio and the risk-free asset can be calculated in a straight forward manner with the simple linear relationship here:

$$\mu_p^e = r_f + (\mu_{p,t} - r_f) \quad (21)$$

Where X is the weight placed in the tangency portfolio and X is the weight placed in the risk free asset. For the purposes of this paper it is the tangency portfolio that will be tested. This approach is consistent with the CAPM as well as the methods developed by Markowitz and Sharpe. For different time periods two tangency portfolios will be constructed using historical statistical estimation and implied volatility. These portfolios will then be evaluated going forward, allowing for the comparison of the different returns (performance). Because the inputs to the algorithm will be different, then the tangency portfolios will be different. However, the important question that can be addressed will be how much different they are and how much better or worse either is.

As mentioned earlier, the only difference in the construction of any given portfolio will be the differing values of volatility. The values of the means, correlation, and risk free rates will be the same for both; however, the former will be calculated using a rolling

twenty-four month historical average. One important nuance is that the Markowitz algorithm uses a variance-covariance matrix, and this matrix will contain differing values where the historical variances will be replaced with implied volatility directly and the covariance will be replaced using the following calculation:

$$\sigma_{ij(IV)} = \rho_{ij(\text{historical})} \sigma_{j(IV)} \sigma_{i(IV)} \quad (22)$$

The covariance values for the implied volatility matrix will be a product of the two implied volatility values with the historical correlation value. Because the portfolios will use this rolling average, the first two years of data will be used as a warm-up period and will not be evaluated in the same context. So from January 2005 to December 2006 all the data will be used to develop the first set of parameters for testing the first month of 2007. It is also important to point out that work has been done on developing future estimates of the other parameters (i.e. means) such as with Frost and Savarino [24] who employ a Bayesian method for better estimating parameters. However, for the purpose of this paper, the historical estimates will be employed though they contain known estimation error. Because the parameters are consistent in both portfolios, the impact of this error will be the same on each and the differences can be attributed to the differing variance estimates. For the portfolio test, a one month holding period will be used. After each holding period, the portfolios will be rebalanced and then evaluated again using the updated parameters (rolling twenty-four months and updated IV). The result will be, using a holding period of one month, forty-eight different unique portfolios for both historical estimation and implied volatility. Then these will be evaluated for both their returns and their Sharpe ratios (the measure of risk for selecting the tangency portfolio). The formal hypothesis test is the evaluation of comparing the actual returns of various portfolios with securities selected using the Markowitz algorithm utilizing the volatility parameter estimations from both historical simulation as well as implied volatility. This hypothesis test will be conducted over the various holding periods from a given data set and the following null hypothesis will be tested

$$r_{IV} = r_{SE} \quad (23)$$

Where r_{IV} is the return of a portfolio over the data set using IV and r_{SE} is the return of a portfolio over the data set using the historical statistical estimation. More formally, the hypothesis test will focus on the statistical difference between the averages of the returns over the various holding periods using the two different methods. The null and alternative are as follows:

$$H_0 : \mu_{IV} - \mu_{SE} = 0 \quad (24)$$

$$H_1 : \mu_{IV} - \mu_{SE} \neq 0 \quad (25)$$

$$\text{significance level: } \alpha = .05 \quad (26)$$

Secondly the risk of either method will be evaluated by comparing the Sharpe ratio from the results of the test. This ratio takes the form:

$$SharpeRatio = \frac{\mu - r_f}{\sigma_p} \tag{27}$$

These two tests will give an understanding of how these two parameter estimates compare to each other in their predictive abilities. We will also be able to see how they each react to the financial crisis and recovery that takes place following the warm-up periods. Another point of analysis will be the comparison of the weights in each of the portfolios over time. As stated earlier, we should see a shift away from the stock assets prior to the crash if the parameter has good predictive abilities. In the results section the weights in each portfolio will be plotted as a time series and the changes will be discussed. Figure 5 is an example of how the two portfolios are selected and then compared, the chart shows the efficient frontiers given the two selection methods and their respective tangency portfolios in the month of January 2007 for a one month holding period: Moving forward one month the actual returns for these two portfolios are given in Table 3.

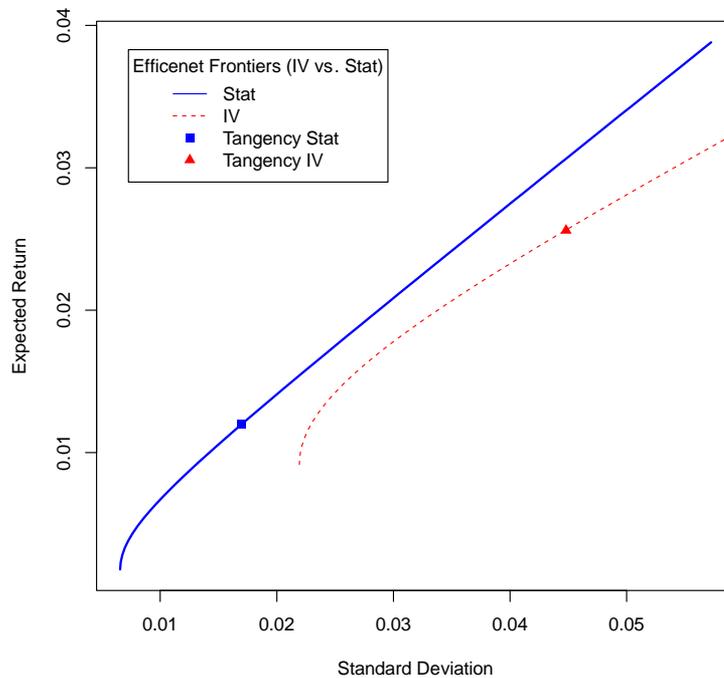


Figure 5: Comparison of Efficient Frontiers

Table 3: Actual Portfolio Performance

Statistical Tangency Portfolio Return	3.93%
Implied Volatility Tangency Portfolio Return	8.79%

5 Results

The first step in this process is to consolidate the implied volatility values referenced from the data and methodology sections. Figure 6 shows the moving values for each portfolio component in a time series over the tested horizon: Two major features of Figure 6 are that it shows how dramatic the options market reacted to the financial crisis through the real estate ETF (VNQ).

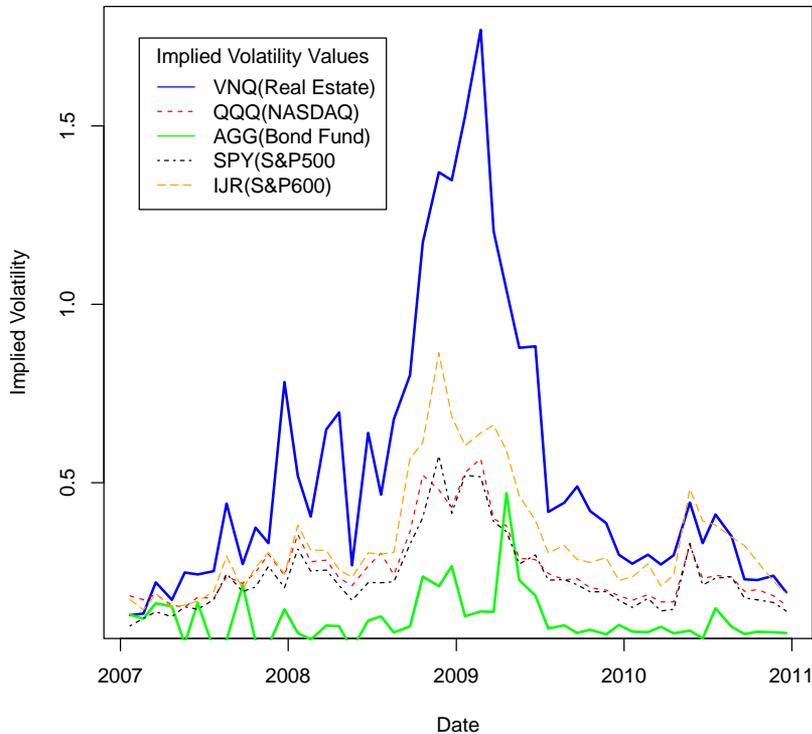


Figure 6: Implied Volatility Values by Portfolio Component

It is clear that this fund has the largest values during and following the crisis in 2008-09. Also it is important to note that the bond fund enjoys fairly consistently low relative implied volatility compared to the other funds save for one spike in mid 2009. From this result we can expect the algorithms to likely disfavor real estate and favor bonds depending on the mean values. It is also worthwhile to view the implied volatilities against the behavior of the underlying fund. The following set of graphs shows the implied volatility values plotted against standard deviation is given to show the difference between it and the IV value in Figures 7 through 8.

In Figure 7 we can see that for SPY (the S&P 500 ETF), the implied volatility appears to be negatively correlated with returns and that its value does not seem to predict, but instead follow, the returns of the fund.

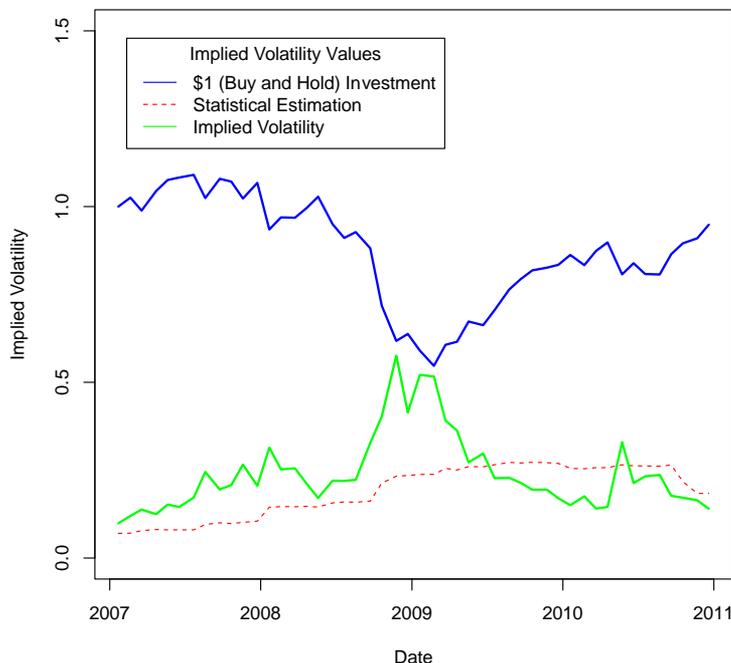


Figure 7: SPY Standard Deviation Estimation Analysis

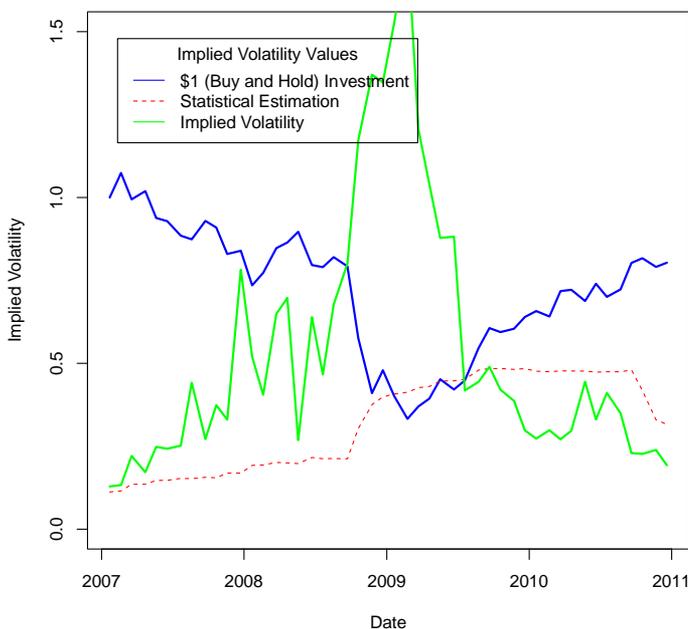


Figure 8: VNQ Standard Deviation Estimation Analysis

With the VNQ (real estate fund) in Figure 8 we see the dramatic reaction that was evident in the consolidated graph. There does however appear to be significant increase in implied volatility prior to the significant crash in late 2008. However, after this, the implied volatility sky rockets and does not return to lower levels until 2010. We can see that during this time this fund is actually an excellent investment recovering most of its value.

In general from the graphs demonstrate that implied volatility is closely related to the underlying returns and they seem, visually, to be more reactionary than predictive in nature. While the returns of the constructed portfolios will be the better test, it is appropriate to explore this observation further. To do this we look at the correlation between the changes in implied volatility and the actual returns during the tested time period which are given in Table 4.

Table 4: Correlation of Changes in IV to Returns

Asset	Correlation
SPY	-.730
QQQ	-.804
AGG	-.143
IJR	-.817
VNQ	-.730

Table IV shows numerically what the graphs suggested. There is a strong statistically significant negative correlation between the implied volatility values and the actual returns. From this table it is suggested that occurring simultaneously the greater the loss the higher we can expect implied volatility to be. It is also worth noting that there is a much lower correlation associated with the bond fund. The reasons for this are not clear other than the funds actual returns are consistent over time and the implied volatility varies greatly, which may in part be due to its relative illiquidity. It does also seem that implied volatility tends to overreact to events and takes a long period to return to values consistent with historical estimation. However, this may be an observation only consistent with this data set. If this is the case (that options are over reactionary), we will expect to see portfolios which hold unfavorable or favorable views for too long and then have an adverse impact on the portfolios performance. With these observations, we can now turn to the question of, how does this information fair when applied to portfolio selection?

Evaluation of the portfolios will take place through the visual interpretation of the data as well as formal statistical hypothesis testing. The first place to look is at how the weights in the portfolios vary over time. Because the portfolios are rebalanced every month, we can get an indication of how they are affected by the events or sentiment that occurs over that period. Presumably, the statistical estimates will not change drastically from period to period as they are estimates drawn from a twenty-four month moving average. However, if there are singular large events that can skew an average, we will see this reflected in the weights accordingly. We can see how the rolling weights behave in Figure 9.

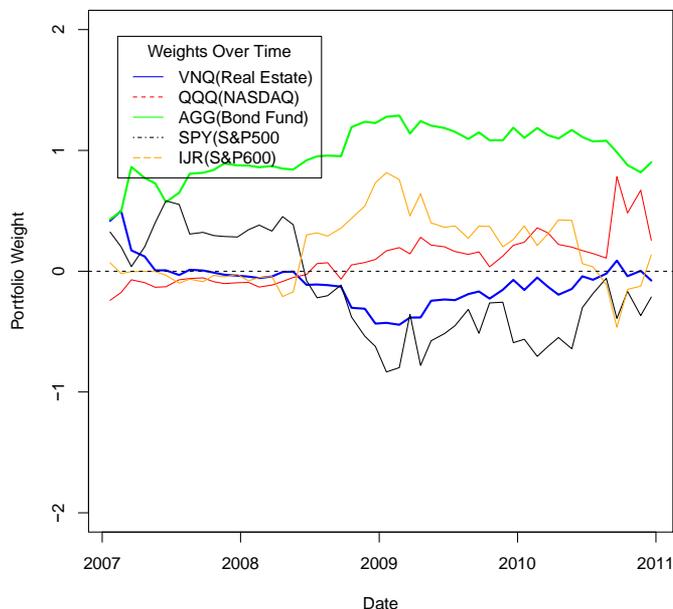


Figure 9: Historical Statistical Portfolio Weights of Components

As we can see, there are very little major moves in any of the component weights. However, they do follow certain trends over time. It is clear that the major trends change dramatically during the financial crises at the end of 2008 and beginning of 2009. This is the period when we observe the large events which change the algorithm significantly. During this time we see a very favorable outlook on AGG (the bond fund) and negative outlooks on SPY and VNQ (S&P 500 and real estate respectively). This observation is consistent with what we would expect given the events of the time period. Another interesting event is the relatively favorable outlook on IJR and QQQ (small cap and NASDAQ respectively) both of which have greater volatility than SPY. While we do see some jumps towards the end of the time series in all securities, this is likely due to the changes in averages as the significant events in 2008 are no longer included. Generally speaking, this series of portfolios reacts with consistent variation based on the events within the given

time series. The weights in the portfolios using the implied volatility are not as straight forward to predict. In general, we should see some correlation as much of the information used to derive the weights is the same as the historical statistical approach. However, as we have seen earlier, implied volatility estimates can vary greatly, and are often much greater than the averages. We are interested to see if the weights in the portfolios behave in a way that is consistent with a prediction of future outcomes. Figure 10 shows the rolling weights of the implied volatility based portfolios.

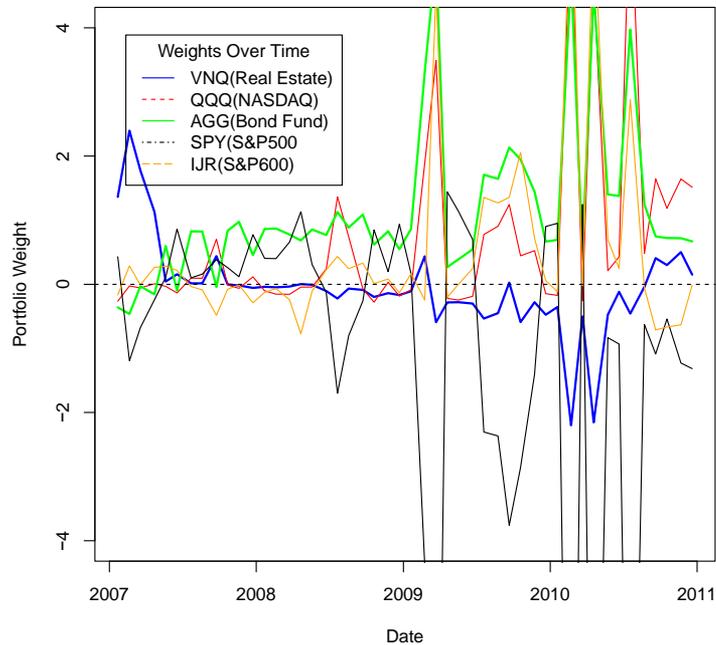


Figure 10: Implied Volatility Portfolio Weights of Components

The first and most significant observation is the dramatic reaction to the crash. We see a significant peak in weight regarding the bond fund. Interestingly this peak is not persistent and only lasts for a few periods in the first part of 2009. While we do see trends in parts of the weights they are not as consistent and display much more variation. For the first three years of the time series the weights of most of the components are at or below zero and the portfolio is comprised mostly of AGG (the bond fund). This trend breaks in the last few years where again we see a favorable outlook on IJR and QQQ. One more interesting note is that in the first few months VNQ (real estate) is strongly preferred which is consistent with market sentiment. However, it quickly loses its weighting and becomes a short position in nearly the rest of the time series. The last significant observation is the dramatic difference in weights with this method. We see weighted values in to the multiple hundreds of a percent, which is quite a large amount of variation when compared to the historical method and not realistic for an average investor.

Now that we have reviewed the weights, we can turn our attention to the actual performance of the approaches. Figure 11 depicts the monthly returns of each approach plotted against each other.

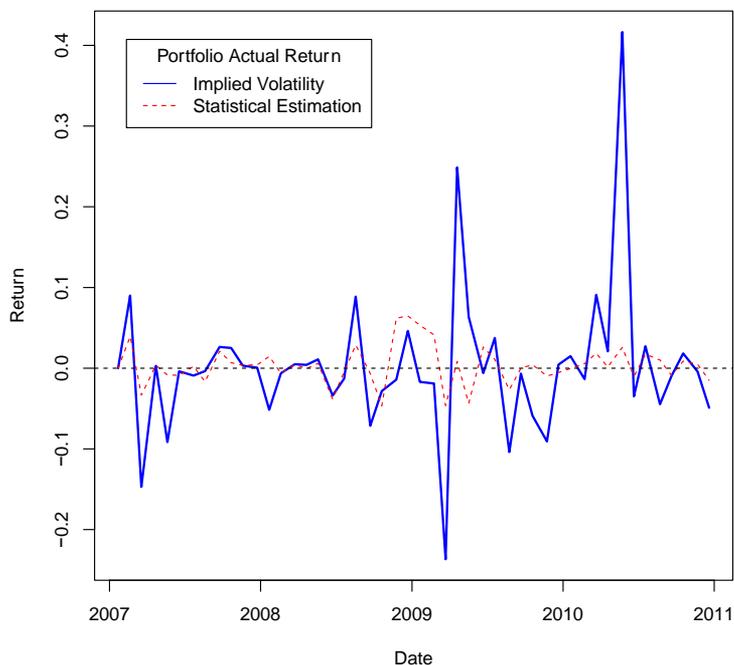


Figure 11: Portfolio Actual Returns

As we can see the two returns are fairly consistent in performance. However, it is clear in the more extreme events the implied volatility portfolios have a much greater variation. The standard deviation of the returns for the two approaches is significant in that statistical approach yielded 2.48% and the implied volatility approach is 9.3%. It is also important to look at the riskiness associated with each approach represented by their respective Sharpe’s slope:

Table 5: Average Performance of Portfolios

Portfolio Performance	Stat	IV
Mean	.34%	.16%
SD	2.48%	9.3%
Sharpe Slope	13.87%	1.76%

As we can see the slope of the statistical estimate is far greater than the implied volatility which also speaks to its lower variation. It is significant that the average return for the implied volatility approach is half that of the historical approach, and reflects a poor result. We see also that there does not seem to be any clear advantage to the implied volatility portfolios and it may actually be a detriment. In fact, the use of implied volatility seems to only result in greater variation and no greater return. One clear example of this is the large loss in early 2009 (approximately seventeen percent) is due to the overreaction stated earlier of the implied volatility estimates. During this period, the market began its recovery but the IV approach was still heavily shorting assets which were hit the hardest during the crash. It would seem from this view that the implied volatility has more of a worse case of bias than do the historically estimated values. It is also valuable to look at

the total returns over time. Similar to earlier graphs of the individual components, Figure 12 shows the growth of one dollar invested in each approach over the entire time series.

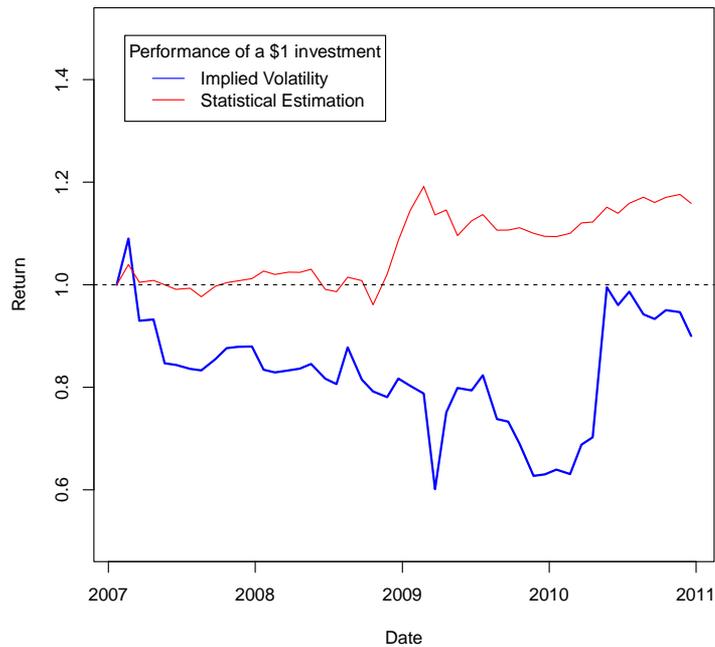


Figure 12: Performance of \$1 Invested in Each Strategy

From this graph we can see that for the first few years there seems to be little difference between the two approaches (save the greater initial variation). The major difference is that the implied volatility estimates remained high in 2008-09 and lead to poor weighting of the portfolio to take advantage of any recovery. In fact, the opposite happens: There is a significant loss due to short positions in recovering assets.

6 Summary

It seems that the tendency for implied volatility to remain high after significant events leads to losses that more than offset any gains made prior. It is worth pointing out, however, that from late 2009 onward implied volatility does outperform historical statistical estimation. Even though during the entire tested time period it is the opposite. It is not clear if either approach is statistically significantly better than the other. In order to test the differences in the approaches a paired t test was conducted with the results detailed in Figure 5.

Table 6: Test of Statistical Significance

Paired t-Test	
Average Difference	35.00
Alpha	35.00
Test Stat	2.01
n	47
S	0.0857
Paired t-Test Result	0.14

As we can see from the table, the result is that we fail to reject the hypothesis that the two approaches are the same. Basically there is no statistical basis to say that one approach is definitively better than the other. This is a surprising result as much of the literature evaluates implied volatility as a better estimate than the historical statistical method. It seems however, when put into practice, its faults (over reaction and high variation) do not outweigh its benefits. We can say that at times implied volatility is better and at other times it is not, but which times are which are not clear.

References

- [1] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley, 1959.
- [2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, **81**, pp. 637–654, 1973.
- [3] S. Neftci, *Principals of Financial Engineering*. London: Academic Press, 2008.
- [4] S. Manaster and G. Koehler, "The calculation of implied variances from the black-scholes model: A note," *The Journal of Finance*, **37**(1), pp. 227–230, 1982.
- [5] J. S. Doran, D. R. Peterson, and B. C. Tarrant, "Is there information in the volatility skew?" *Journal of Futures Markets*, **27**(10), pp. 921–959, 2007.
- [6] M. Fengler, *Semi Parametric Modeling of Implied Volatility*. Berlin: Springer, 2004.
- [7] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007, 1982.
- [8] S.-H. Poon and C. W. Granger, "Forecasting volatility in financial markets: A review," *Journal of Economic Literature*, **41**(2), pp. 478–539, 2003.
- [9] L. H. A. and J. Rendelman, "Standard deviations of stock price ratios implied in options prices," *Journal of Finance*, **31**, pp. 369–381, 1976.
- [10] D. Chiras and S. Manaster, "The information content of option prices and a test for market efficiency," *Journal of Financial Economics*, **6**, pp. 213–234, 1978.
- [11] R. Schmalensee and T. R. R., "Common stock volatility expectations implied by option premia," *Journal of Finance*, **33**, pp. 129–147, 1978.
- [12] J. Shu and J. Zhang, "The relationship between implied and realized volatility of sp 500 index," *Wilmott magazine*, pp. 83–91, January 2003.
- [13] J. Fleming, "The quality of market volatility forecasts implied by S&P 100 index option prices," *Journal of Empirical Finance*, **5**(4), pp. 317–345, 1998.
- [14] B. J. Blair, S.-H. Poon, and S. J. Taylor, "Forecasting s&p 100 volatility: the incremental information content of implied volatilities and high-frequency index

- returns,” in *Handbook of Quantitative Finance and Risk Management*. Springer, 2010, pp. 1333–1344.
- [15] L. Canina and S. Figlewski, “The informational content of implied volatility,” *Review of Financial studies*, **6**(3), pp. 659–681, 1993.
- [16] Y. Xing, X. Zhang, and R. Zhao, “What does the individual option volatility smirk tell us about future equity returns?” *Journal of Financial and Quantitative Analysis*, **45**(3), p. 641, 2010.
- [17] R. E. Whaley, “The investor fear gauge,” *The Journal of Portfolio Management*, **26**(3), pp. 12–17, 2000.
- [18] Corrado and Miller, “The forecast quality of CBOE implied volatility indexes,” *The Journal of Futures Markets*, **25**(4), pp. 339–373, 2005.
- [19] R. E. Whaley, “Derivatives on market volatility: Hedging tools long overdue,” *The journal of Derivatives*, **1**(1), pp. 71–84, 1993.
- [20] P. S. Banerjee, J. S. Doran, and D. R. Peterson, “Implied volatility and future portfolio returns,” *Journal of Banking & Finance*, **31**(10), pp. 3183–3199, 2007.
- [21] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, “Early-warning signals for critical transitions,” *Nature*, **461**(7260), pp. 53–59, 2009.
- [22] D. P. Simon, “The NASDAQ volatility index during and after the bubble,” *The Journal of Derivatives*, **11**(2), pp. 9–24, 2003.
- [23] G. Kaplanski and H. Levy, “Sentiment and stock prices: The case of aviation disasters,” *Journal of Financial Economics*, **95**(2), pp. 174–201, 2010.
- [24] P. A. Frost and J. E. Savarino, “An empirical bayes approach to efficient portfolio selection,” *Journal of Financial and Quantitative Analysis*, **21**(3), pp. 293–305, 1986.