THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF FINANCE

A TEST OF THE PREDICTIVE POWER OF VOLATILITY FORECASTING MODELS

CHRISTIAN BINDER
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Finance and Economics
with honors in Finance

Reviewed and approved* by the following:

Lou Gattis
Clinical Professor of Finance
Thesis Supervisor

Brian Davis
Clinical Assistant Professor of Finance
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

In this paper, a quantitative evaluation and ranking of commonly used volatility forecasting models and metrics is presented. The models and metrics included are Generalized Autoregressive Heteroscedasticity (GARCH), Exponentially-Weighted Moving Average (EWMA), Implied Volatility, long-term historical mean volatility, and a moving average volatility. The metrics were tested on a randomly-selected sample of US equities, international equity indices, currencies, and commodities on both historical data and in real-time, making this study unique. Models are ranked and compared via various statistical loss functions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Many thanks are given to Dr. Lou Gattis for his unfaltering guidance, infinite patience, and thoughtful counseling throughout the entirety of the thesis process. I would also like to thank Dr. David Haushalter for his suggestions and help throughout the writing process. Final thanks are due to Dr. Brian Davis for his constant encouragement and support.

Chapter 1

Introduction

In this paper, a quantitative evaluation and ranking of commonly used volatility forecasting models and metrics is presented. The models and metrics included are Generalized Autoregressive Heteroscedasticity (GARCH), Exponentially-Weighted Moving Average (EWMA), Implied Volatility, long-term historical mean volatility, and a moving average volatility. The metrics were tested on a randomly-selected sample of US equities, international equity indices, currencies, and commodities on both historical data and in real-time, making this study unique. Models are ranked and compared via various statistical loss functions.

In the late 80s and early 90s, finance and economics journals exploded with literature regarding volatility forecasting. Following the Black Monday crash of October 1987, in which there was a five-standard-deviation change in the value of the overall market instantaneously (an event which should occur approximately only every five billion years), substantial research was devoted towards finding empirical evidence explaining large swings in asset prices, as well as how to forecast these swings. The research has yielded numerous methods of measuring and forecasting volatility. Forecasting volatility is an important function of academics and finance professionals from day-traders to portfolio managers and thus has a wide range of applications and consequences. For example, future volatility is the most important input in the Black-Scholes Option Pricing Model (as the other inputs to the model are exogenous/given); in this instance, knowing which volatility forecasting model is “best” would allow traders to price options more accurately and with more certainty. Volatility also has important applications to

Value-at-Risk (VaR) projections, which is typically done for a longer time horizon, and thus knowing how to best forecast volatility would allow portfolio managers to more efficiently manage their VaR.

Chapter 2 contains a review of the literature both theoretical and empirical on the subject of volatility forecasting. Chapter 3 goes into further detail about the statistical properties and features of the models which are tested, and explains why these models and metrics were chosen for this study. Chapter 4 describes the methodology used for analyzing and ranking the models. Chapter 5 analyzes and interprets the data and presents a quantitative ranking of the tested models and metrics. Chapter 6 summarizes the main points of this paper and presents a few suggestions for further research.

Chapter 2

Literature Review

In the seminal paper *The Pricing of Options and Corporate Liabilities* (1973), Fischer Black and Myron Scholes laid the foundation for modern option pricing theory when they presented the first quantitative method of pricing vanilla European-style options based on five underlying variables: the current stock price, the strike price, the time to maturity, the interest rate, and most importantly the expected volatility. While the first four variables are known and exogenous to the Black Scholes Model, the future volatility over the expected life of the option is unknown and must be forecasted; in other words, volatility is an endogenous input to the Black Scholes Model and thus will have the largest impact on determining the price of a given option. This means practitioners and traders have a vested interest in obtaining accurate measures of future volatility over a given time horizon, as more accurate forecasts of volatility will enable traders to determine whether a given option is currently over or under priced in the market.

The GARCH-Family Models

After the introduction of the Black Scholes model, practitioners often relied on simple rolling one-month standard deviations for the next month's volatility, thereby applying equal weights to all individual observations of volatility and leaving out observations older than one month. In 1982, Robert Engle published a paper in *Econometrica* which introduced a new method of forecasting volatility, which essentially let the weights applied to each observation of volatility be parameters which need to be estimated. Engle argued that previous econometric models relied on faulty assumptions about the behavior of volatility, thus inhibiting those models' accuracy and predictive power. Engle went on to say that error terms in time series processes are not necessarily homoscedastic. He further argued that due the empirically proven phenomena of autocorrelation in volatility (also known as volatility clustering; Mandelbrot, 1963) present in financial time series that a new model should be built. In this paper, Engle introduced an autoregressive conditional heteroscedastic (ARCH) process and model. In these processes, recent volatility informs one-period ahead forecasts of volatility. Of important note is that a feature of this model is that Engle allowed for unconditional (long-term) variance to stay constant while allowing for changes in conditional variance to change over time.

In his 1986 paper titled "Generalized Autoregressive Conditional Heteroscedasticity," Tim Bollerslev generalized Engle's ARCH process to a GARCH process. Bollerslev claimed that there was an issue with Engle's simple ARCH process as it has an arbitrary strictly linear declining lag structure in the conditional variance equation which allows only for past in-sample variances. The GARCH, however, process allows for lagged conditional variances to enter the process as well. In simpler terms, this means that although the GARCH model is intended to forecast one-period ahead, it can produce a two-period ahead forecast by considering the initial first-period GARCH forecast (and so on for n-periods), thus allowing for longer time horizon forecasts than the initial ARCH model. As forecasts are made further into the future, the GARCH predicted volatility will converge toward the long-run unconditional variance. These properties of the model, in addition to its inherent simplicity, have made GARCH one of the most commonly used volatility forecasting models today.

Following up on the original GARCH model, numerous academics proposed extensions of the GARCH model which attempted to incorporate factors which accounted for additional properties of volatility, resulting in an alphabet soup of GARCH models. While explaining all of these GARCH-family models and their properties is beyond the scope of this paper, there is one that is particularly worth mentioning. One of the most commonly used GARCH extensions is the exponential-GARCH (EGARCH) model introduced by David Nelson in 1991. Nelson cited Fischer Black's 1976 paper which found a negative correlation between stock returns and changes in the returns volatility. Black demonstrated that volatility rises in response to bad news and falls in response to good news; Nelson points out that this implies an asymmetric distribution of changes in conditional volatility and that GARCH fails to account for this empirical observation. He claims that a GARCH process which also accounts for asymmetry in residuals should more accurately forecast volatility. Nelson develops a model which accounts for this and which has a few other unique properties as well. The EGARCH functions similarly to the canonical GARCH model. The EGARCH model, while more complex than the orginal GARCH model, is still simple enough for everyday use and is commonly used amongst practitioners to forecast volatility and price options.

Implied Volatility

Another way of measuring future volatility is through implied volatility. As previously mentioned, the Black Scholes model is typically used to determine the price of an option, and the model takes volatility as an input. However, the Black Scholes model can be rewritten, taking the market price of an option as given, to solve instead for the volatility implied by the market. In other words, this is the measure of volatility, that when put into the Black Scholes Model, will give as an output the current market price of the option.

The reason implied volatility can be used as a forecast of future volatility is that it is the volatility that is expected over the remainder of the life of the given option. In this sense, it is also an extension of the Efficient Market Hypothesis. Because the EMH would argue that the market price of the option reflects all available information of that option, the implied volatility consists of all available information with respect to the volatility of that option for the remainder of its life. In other words, implied volatility theoretically consists of all available forecasts of future volatility in the market which are used to price the option, including forecasts as simple as historical volatility and those as complicated as the GARCH-family models. Therefore, implied volatility is also considered "the market's view" of volatility.

The main issue with using implied volatility as a forecast of future volatility is that there may be multiple implied volatilities on the same underlying asset. Options on the same underlying asset which have different strike prices may offer different implied volatilities. Similarly, options with different times to maturity and identical strike prices will offer different implied volatilities as well (however, this makes more sense as the volatility can be expected to be different for differing periods of time). Naturally, these two features can be combined to create what is known as an implied volatility surface in three-dimensional space, with time to maturity on the x-axis, implied volatility on the y-axis, and strike prices on the z-axis. If the Black Scholes Model could accurately determine the price of all options, such a surface would be flat. However, this is not the case since we know some of the assumptions underlying Black Scholes are empirically untrue (explained in the next paragraph). In an attempt to simplify this, we will use the time-scaled average of the implied volatilities of at-the-money call and put options for each randomly selected stock as a forecast for future volatility.

A second weaknesses of using implied volatility as a forecast of future volatility (and therefore a weakness of the papers which empirically test its forecasting accuracy) is that volatility implied by the Black Scholes Model does not necessarily lend itself well to describing the true nature of volatility or option prices. In other words, using volatility implied by the Black Scholes Model implies that all the assumptions underlying Black Scholes are true. However, we know empirically that this simply isn't the case. For example, papers by both Hull and White (1987) and Heston (1993) demonstrate that price patterns of options exhibit stochastic (non-constant) volatility; each of these papers present new stochastic volatility models for pricing options. One could derive implied volatilities from stochastic volatility option pricing models, and thus we should be more precise in specifying which implied volatility we are measuring. However, deriving implied volatilities from stochastic volatility models is incredibly mathematically complex and is not easily done by the average practitioner or trader.[1]

Since stochastic volatility models, such as the Heston Model, more accurately reflect empirically-proven characteristics of volatility, one would reasonably expect them to forecast volatility more accurately than simple Black Scholes implied volatility. However, due to the mathematical complexity involved in these models, I will limit our tests to Black Scholes implied volatility while conceding that further research should seek to test stochastic volatility models' implied volatilities.

---

[1] Those truly interested in seeing derivations of implied volatility from stochastic volatility models should consult "Computing the Implied Volatility in Stochastic Volatility Models" (Berestycki, 2004), and "Asymptotic Formulae for Implied Volatility in the Heston Model" (Forde, 2010).

Empirical Studies

After the proliferation of the GARCH family models in the late 80s and early 90s, theoretical literature on volatility models slowed down. For the most part, the literature generally shifted from theory towards empirical tests of the models' predictive accuracy. The first such empirical paper was published in 1990 by Pagan and Schwert. In this paper, the authors laid the groundwork for all empirical tests of volatility models to follow. The authors tested the predictive accuracy of a selection of volatility models on monthly stock return data from the period 1835-1925. The authors include GARCH and EGARCH in their test, in addition to nonparametric Fourier Form model and a Markov regime-switching model. After fitting their models to and testing on in-sample data, the authors test the same set of models against monthly return data using an out-of-sample forecast for the period of 1925-1937. The in-sample tests and out-of-sample tests reach contradictory conclusions: the nonparametric models (such as the Fourier Form and Markov Chain model) performed better in-sample, while the parametric models (GARCH and EGARCH) performed better on the out-of-sample tests. From this, the authors argued that the parametric and nonparametric schools of volatility forecasting needed to be unified to have the most predictive accuracy; this is in part what led to the further development of more mathematically complex GARCH-family models in the early 90s.

The largest issue with the Pagan and Schwert paper is the reliability of the data. How reliable is monthly return data from the 19th Century? Pagan and Schwert do not indicate their source for this data, further calling into question the reliability of the data and the authenticity of their results. The age of the data also calls into light the issue of liquidity. With the proliferation of high frequency trading, financial markets are now more liquid than ever and react to news significantly faster than they had in the 19th and early 20th Century. Different levels of liquidity may imply different levels of volatility (the causality can also run in the reverse direction) and thus the authors' results may not as valid in today's markets. Another issue with this paper is that the authors tested volatility forecasting models which are almost never used by practitioners, namely the Fourier Form models and the Markov Chain regime-switching models. Although GARCH and EGARCH were included in the test, the study would have benefited from including more traditional methods of forecasting volatility, such as simply using historical volatility.

In 1993, Theodore Day and Craig Lewis published a paper which tested the accuracy of GARCH-family models and implied volatilities on daily prices for call options on crude oil futures. Like Pagan and Schwert before them, Day and Craig test the models on both in-sample and out-of-sample data. However, the authors expanded on what Pagan and Schwert did by testing the models against different time horizons. The authors test the models for both the shortest time to maturity and longest time to maturity options; this is important because certain models may be better at forecasting on shorter time horizons, while other models may be more accurate at forecasting for longer time periods. The authors conclude that implied volatility performs better on average to forecast the volatility of crude oil futures, both on the shorter and longer time horizon. One of the largest issues with this paper, however, is that the authors chose to test their models on a time series which experienced a large exogenous shock to the oil markets: the US' invasion of Kuwait and the Gulf War from 1990-1991. During this time, oil prices were extremely volatile, rising from about $31 before the invasion to a peak of about $72 during the invasion. The issue in doing this is that the accuracy of these results are only valid for periods of heightened volatility for a single asset class and thus may not be applicable in times of lower volatility or for other asset classes.

Other tests have examined volatility forecasting metrics individually by simply comparing them to realized volatility, an approach more in-line with the methodology of this paper.

One of the most widely-cited empirical papers testing volatility models is “A Comparison of Volatility Models: Does Anything Beat GARCH (1, 1)” (Hansen & Lunde, 2005). In this paper, the authors attempt to compare out-of-sample forecast accuracy of 330 volatility models, including almost all the GARCH-family models with lags different than the traditional (1, 1), on intra-day exchange rate data and a single stock price. The authors chose to test the models on high-frequency intra-day data due to an argument put forth by Andersen and Bollerslev in 1998 that high-frequency data will yield improved forecast performance (one would think this is a trivial observation as a larger sample should result in a better fit). The authors then rank the models on seven different criteria, including MSE and mean absolute deviation. The authors found that no other models significantly outperform the GARCH (1, 1) benchmark model on the exchange rate data. However, the results were less robust for the equity data, as some models yielded p-values as high as 0.10, which the authors claim signals that a better forecasting model should exist. Of equal importance is that each model tested ranks differently by each loss function tested (in other words a given model may have one of the lowest MSEs but a very high Akaike Information Criterion [AIC]), emphasizing the importance of choosing the best loss function on which to rank models. Unfortunately, defining what is the “best” loss function is still a measure of intense debate within statistics and econometrics. Hansen and Lunde concede the faults of their paper in their final chapter. The authors admit that different models may perform differently on different assets, and that further tests should be done on longer-term volatility using lower frequency data than intra-day. This paper attempts to address a few of the concerns of Hansen and Lunde by testing a few volatility models' longer-term forecast performance on daily returns data for a broader sample of equities.

Over the past decade, empirical work on volatility forecasting has slowed. The research that has been done has shifted towards using high-frequency data to predict intra-day volatility. Papers such as "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data" (Zhang et al, 2005) and "A Fourier Transform Method for Nonparametric Estimation of Multivariate Volatility" (Mallavian and Mancino, 2009) have laid the theoretical groundwork for modeling such high frequency data, and developed a model known as the Integrated Volatility via Fourier Transformation (IVFT) model. Empirical papers such as "Return Distributions and Volatility Forecasting in Metal Futures Markets: Evidence from Gold, Silver, and Copper" (Khalifa et. al, 2010) have tested these high frequency models. Although this is clearly of interest to the high-frequency trader academics, we will ignore these models as they are needlessly mathematically complex. We will simply concede that future research should test these models against those tested in this paper.

However, there have been a few recent studies which followed the traditional method of testing simpler volatility models. One paper published in 2013 tested univariate and multivariate GARCH models on oil, natural gas, and electricity data (Efimova & Serletis, 2014). Like others before them, the authors found that univariate GARCH models perform better than their mathematically-complex multivariate peers.

In late December of 2015, Stephen Marra, a Portfolio Manager at Lazard Asset Management, published a report on forecasting volatility that is similar in vein to this paper. In the report, titled "Predicting Volatility," Marra weighs the pros and cons of the very same volatility models which will be tested in this paper, and uses a MSE ranking of their predictive accuracy. Based on his tests, ARMA and GARCH produce the lowest MSE, while the simple historical mean had the highest MSE. Unfortunately, this paper was already in the works when this report was published. However, there are quite a few issues with Marra's report. First, Marra does not offer any explanation to his methodology outside of mentioning that the models were tested out-of-sample on monthly returns for the period of 1988 to 2015; we do not know if this was on simply S&P 500 returns, multiple equities or indices, or currencies. We also do not know if he used a rolling out-of-sample forecast, how many periods ahead he forecasted (one, two, ten, or one hundred), or the size of his out-of-sample relative to in-sample data. Because of the lack of description of methodology, the report lacks credibility. In this paper, methodology will be explicitly detailed so that the results are easily replicable. Additionally, it does not seem as if Marra tested his selected metrics in real-time on live data. The methodology of this paper will include such tests, as this will reduce the issue of overfitting with out-of-sample predictions, in addition to being the most comparable to what traders do when forecasting volatility for pricing options.

The biggest problem with the recent literature on volatility forecasting is too much emphasis has been put on developing increasingly complex and obtuse models which attempt to "better" explain the behavior of volatility relative to the amount of effort that has been put into empirical studies of various models' predictive power. Furthermore, the empirical studies that have been done are flawed as described above, and almost all of them rely on out-of-sample methodology which is prone to overfitting data. Since the introduction of the GARCH model, volatility models have become so complex that one needs a graduate-level understanding of mathematics and statistics to have the ability to use these models. The implication of this is that these newer models are too complex for the average trader to use in forecasting volatility and are accessible only to academics who have little practical use for such models.

This paper's goal to present an easy to understand comparison of the most widely-used and most broadly applicable models of volatility forecasting, with the hypothesis that GARCH (1, 1) will be the best performer. We will offer forecasts and loss-function rankings for different time periods, as some models may be better on shorter time horizons while others are better at forecasting over a longer period. This will enable readers to choose the model most appropriate for their needs.

Chapter 3

Overview of the Properties of the Tested Forecasting Models

This chapter will provide a detailed mathematical description of the statistical metrics used throughout this paper.

Using similar notation as Hull in *Options, Futures, and Other Derivatives, Eighth Edition* (2009), we will define returns for an individual asset, S, from time t-1 to time t as

$$r_t = \ln(\frac{S_t}{S_{t-1}})$$

Therefore, the mean return over t periods of an asset S can be expressed as

$$\bar{r} = \frac{1}{t}\sum_{i=1}^{t} r_{t-i}$$

The sample variance of this asset S, using a total of t observations, can be expressed as

$$\sigma_t^2 = \frac{1}{t-1}\sum_{i=1}^{t}(r_{t-i} - \bar{r})^2$$

It follows that the volatility of an asset at time t is simply the square root of this variance. We will define this as "Long Term Historical Average Volatility" for the remainder of the paper.

An n-day moving average volatility will be defined as a subsample of the long-run historical average volatility for t periods taken over the last n days, or

$$\sigma_n = \sqrt{\frac{\sum_{i=t-n}^{t}(r_i - \bar{r})}{n}}$$

This paper will evaluate a 50 Day Moving Average and a 200 Day Moving Average, as these are the two of the most frequently used time periods in practice.

Alternatively, as we will be monitoring the daily level of volatility in this paper, it is often helpful to write these equations as follows:

$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}}$$

for the return of an asset, and

$$\sigma_t^2 = \frac{1}{t}\sum_{i=1}^{t} {r_{t-i}}^2$$

for the variance. Again, it follows that the daily volatility of an asset at day t is simply the square root of this variance.

Long Term Historical Average Volatility applies equal weighting to each individual observation of volatility by design, regardless of how long ago that observation occurred. As previously discussed in Chapter 2, the EWMA and ARCH/GARCH models attempt to estimate optimal weights to apply to each observation of volatility. Naturally, each of these models give more recent observations more weight than older observations. We can then define a weighted model of volatility as

$$\sigma_t^2 = \sum_{i=1}^{t} \alpha_i {r_{t-i}}^2$$

where the term α_t is a strictly positive weight given to the observation i periods before. By definition, the weights must sum to 1

$$\sum_{i=1}^{t} \alpha_i = 1$$

Recall our previous definition of Long Term Historical Average Volatility. By squaring this term to get the long term historical average variance, we can define the long term historical average variance, V_L, as the sum of the squared log returns of asset S:

$$V_L = \frac{1}{t}\sum_{i=1}^{t} r_i^2$$

An ARCH process, as defined by Engle (1982), assigns a unique weight, γ, to this long run average variance while applying individual weights to individual observations. The older the observation, the less weight it is given. This can be formally expressed as

$$\sigma_t^2 = \gamma V_L + \sum_{i=1}^{t} \alpha_i {r_{t-i}}^2$$

Traditionally, we let

$$\omega = \gamma V_L$$

so that the ARCH process can be rewritten as

$$\sigma_t^2 = \omega + \sum_{i=1}^{t} \alpha_i {r_{t-i}}^2$$

We can express GARCH (1, 1) as

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

with γ being the weight assigned to the long run historical average variance, α being the weight assigned to the previous period's return, and β being the weight assigned to the long-term average sample variance. By design,

$$\gamma + \alpha + \beta = 1$$

Additionally, GARCH (1, 1) processes always have the constraint that

$$\alpha + \beta < 1$$

where $\alpha\ and\ \beta$ are parameters estimated by finding values for those parameters which maximize the value of the function

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \frac{-r_i^2}{2\sigma_i^2}$$

or, by taking logs, equivalently

$$\sum_{i=1}^{n} -\ln \sigma_i^2 - \frac{r_i^2}{\sigma_i^2}$$

There are two important features of the GARCH (1, 1) process which have been mentioned in Chapter 2. First, the GARCH (1, 1) weights decay exponentially at rate β. We can show this by substituting the previous period's GARCH process in for σ_{t-1}^2

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta(\omega + \alpha r_{t-2}^2 + \beta \sigma_{t-2}^2)$$

Substituting in another period back yields

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta(\omega + \alpha r_{t-2}^2 + \beta(\omega + \alpha r_{t-3}^2 + \beta \sigma_{t-3}^2))$$

Therefore, we can continue this process for n periods. The second important feature of GARCH (1, 1) is that as we forecast further into the future, the GARCH forecast converges towards the long run historical average variance V_L.

Notably, this GARCH model only gives us the next period t volatility forecast. This GARCH model can be rewritten using a similar substitution process as above to produce t+i period forecasts. This can be done as follows:

$$\sigma_{t+i}^2 = \left(V_L + (\alpha + \beta)^i\right) * (\sigma_{t+i-1}^2 - V_L)$$

This is the GARCH forecast we will use in forecasting volatility.

The EWMA Model we use for forecasting is a special case of GARCH (1, 1), where

$$\omega = 0, \beta = \lambda \; and \; \alpha = 1 - \lambda.$$

Therefore, we can write this process as

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2$$

Where

$$0 \leq \lambda \leq 1$$

Using recursive substitution as we did for GARCH for this EWMA process will yield

$$\sigma_t^2 = (1-\lambda)\sum_{i=1}^{n} \lambda^{i-1} r_{n-i}^2 + \lambda^n \sigma_{t-n}^2$$

or equivalently

$$\sigma_t = \sqrt{\sum_{i=1}^{n}\left[\frac{(1-\lambda)\lambda^{i-1}}{\sum_{j=1}^{n}(1-\lambda)\lambda^{j-1}}\right](r_{t-i}-\bar{r})^2}$$

This is the number we will use for the EWMA Volatility for forecasting. We use $\lambda = 0.94$ for all forecasts, which is consistent with both common practice and empirical research compiled by JP Morgan's RiskMetrics software (JP Morgan, 1995).

Chapter 4

Description of Methodology

The previous chapter outlined the main volatility forecasting models which are being tested. For the analysis, these models are tested on a random selection of stocks, equity indices, exchange rates, and commodities. Assets for which GARCH (1, 1) or EWMA could not converge were excluded from the process. A list of all current S&P 500 company tickers as of 1/31/17 was downloaded from FactSet. From this, 50 companies were randomly selected. GARCH and EWMA did not converge for 8 of these companies, resulting in a usable sample of 42 randomly selected companies. A similar process was done for 10 randomly selected major currency exchange rates; five of these exchange rates remained after discarding non-convergent datasets. This was done again for a set of 5 indices, 3 of which did not converge, leaving us with two equity indices (S&P 500 and Nikkei 225). Finally, five commodities were randomly selected, and again three did not converge to a feasible solution, leaving us with Gold and Silver. Overall, this yields us a sample of 51 randomly selected assets for forecasting.

For each of these assets, five years of historical price data from the period 1/27/2012 to 1/27/2017 was downloaded from FactSet. Each model was individually fit to this historical time series for each asset, except implied volatility. Implied volatility was measured as the average of the implied volatilities of at-the-money call and put options with a time to maturity closest to one month from 1/27/2017 from FactSet. Implied volatilities were used only for equities and thus were not tested on currencies, commodities, and the two indices. However, implied volatility was tested on a large enough sample of equities for results to be significant with the caveat that it is only significant for that asset class. Each model was tested in real-time for the month of February as new data came in, just as a trader and/or portfolio manager would forecast volatility in practice. This avoids overfitting and bias in the forecasting models.

Forecasts are specified and tested further in three ways, presenting three distinct time horizons: non-rolling, weekly rolling, and daily rolling. The non-rolling forecasts were done using only the original five years of historical data for each asset. In other words, the models were not re-specified or updated as new data came in during the month of February. This is equivalent to saying the non-rolling set is a one-month forecasted data set on all 51 assets. The models forecast volatility for each trading day in the month of February and March 1st, resulting in 23 individual daily forecasts for each of the six models for each asset, or 138 total forecasts for each asset and 7038 forecasts in total. The non-rolling set is the largest data set.

The weekly-rolling data set was updated at the end of each trading week for a random selection of 10 assets from the 51 randomly selected assets because of computational intensity. These assets are SPY, ADP, DIS, VZ, MET, OXY, USDEUR, EURGBP, Nikkei 225, and Gold. In other words, this set would include five years of historical data plus the first week data from February for the second week's forecasts; the second week's data would be added to produce the third week's forecasts, and so forth. This data set is saying we are forecasting daily volatilities re-specified weekly for a period of one month, with 1380 individual forecast observations total.

Finally, the daily-rolling data set contains the same 10 assets used for the weekly-rolling set. Instead of being tested during the month of February, these daily forecasts were made in real-time for the week of 2/27/17 to 3/3/17. Each model included the original five years of historical data plus the data for the month of February for the daily-rolling set. These models were re-specified at the end of each day incorporating that day's data to produce the next day's volatility forecast. This is the same as saying we are producing daily volatility forecasts for a period of one week, or five daily forecasts across six models for ten assets for a total of 300 individual observations of volatility.

Importantly, the 50 Day Moving Average and 200 Day Moving Average volatility forecasts were rolled daily across all rolling frequencies (otherwise they wouldn't be considered moving averages, which is what we want to test). Implied volatility was not rolled at all because of computational intensity.

Once the forecast periods were over, realized volatility for each individual day for the forecast period was calculated. Despite tedious academic disagreement over how to calculate realized volatility, the simple approach to calculating realized volatility most often used in practice was used. Mathematically, realized volatility for an individual observation was calculated as

$$\sigma_t = \sqrt{\sigma_t^2} = \sqrt{(r_t - \bar{r})^2}$$

where r_t is the asset's return for that day and $\bar{r}$ is the long-run average return for that asset, where the long-run average was also updated in accordance with the rolling frequency.

Each model's accuracy was measured using a variety of statistical loss functions. Since there is no single best loss function, a variety of widely used loss functions was used to assess the models. The loss functions were Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Absolute Value of Mean Forecast Error (Abs MFE), and Log Loss Function (LLF). These are defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{\sigma}_t - \sigma_t)^2$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\sigma_t - \hat{\sigma}_t}{\sigma_t}\right|$$

$$RMSE = \sqrt{MSE}$$

$$Abs\ MFE = \left|\frac{1}{n}\sum_{i=1}^{n}(\sigma_t - \hat{\sigma}_t)\right|$$

$$LLF = \frac{1}{n}\sum_{i=1}^{n}(\ln \sigma_t - \ln \hat{\sigma}_t)^2$$

The closer each of these loss functions are to 0, the more accurate the model. Therefore, a model is defined as "best" for a given loss function and given asset if said model has the loss function closest to 0. Each of these loss functions were individually calculated and then ranked for each model for each asset and for each rolling frequency.

Additionally, once each forecast period was over and the loss functions were calculated for each model and each individual asset, each rolling methodology's data and forecasts were pooled and aggregated into a single series upon which loss functions for the models were also calculated. In other words, each individual asset's forecasts and realized values were concatenated into a single series rendering time irrelevant, and thus there is a total of three aggregated series, one for each rolling frequency. The reason for doing this is it allows us to compile a large enough sample to see which model performed best on average across the loss functions. Although our results will include which models ranked the best via a counting scheme for each individual asset, aggregating the data into a single series allows us to provide more context as opposed to simply saying one model was the most accurate x times.

Unfortunately, the chosen forecast methodology prevents us from performing a Diebold Mariano test of the statistical significance between the loss functions' values for each methodology except for the daily rolling frequency. Diebold Mariano requires a large sample of error terms for each forecast period, where for Diebold Mariano, a forecast period is defined as the total number of periods a head a forecast is made for (e.g. 23 for our non-rolling methodology, 5 for our weekly methodology, and 1 for our daily rolling methodology). In other words, the DM test would only consider of the errors of each t=23 term for our non-rolling frequency and would disregard the errors of the t=1,2,3, … ,22 forecasts. A multivariate Diebold Mariano adaptation is possible in theory; however, it has not been successfully implemented in any statistical software for practical application and empirical work. Therefore, we will simply define a model as best according to a given loss function if it has the value closest to 0 for that loss function as previously mentioned, with the nontrivial caveat that the results have not been tested for statistical significance.

The Diebold Mariano test (DM test) of statistical significance will be used for the daily-rolling frequency methodology's aggregated series. We will use the R programming language and the "dm.test" function from the "forecast" package to perform the Diebold Mariano test on the errors produced from the models, where the errors are the difference between the realized volatility and forecast volatility for a given model. Notably, this function takes as arguments two vectors of error terms (one per model for comparison), the alternative hypothesis to be tested (we will test that the difference in loss function values between models is different from 0), the forecast horizon (which for our daily-rolling methodology is always 1 as we are producing 1-step ahead forecasts), and the power of the loss function, where the power of the loss function is either 1 (as is the case for linear loss functions, such as for MAPE and Abs MFE) or 2 (as is the case for quadratic loss functions such as MSE, RMSE, and LLF). Therefore, the DM test will return equivalent test statistic values for loss functions of equivalent power. The DM test will be performed across all combinations of models and will thus produce two matrices of DM test statistics (one for each power of loss functions). Using the parenthetically aforementioned null hypothesis and a critical value of $\alpha=0.05$, any DM test statistics greater in absolute value than 1.96 will be considered statistically significant.

Chapter 5

Empirical Results and Analysis

In this chapter, we will break down the results by rolling frequency and analyze which model was the best at forecasting volatility over the forecast period. At the end of the chapter, we will also detail the overall market environment for the forecast period to provide broader context in which to evaluate the reliability and/or potential bias of the results.

Non-Rolling Forecast

Table 1 below shows the number of times a model had the lowest value for a given loss function out of the total 51 assets for the non-rolling forecast.

Table 1: Number of Times Lowest Loss Function Value for Individual Assets (Non-Rolling)

	GARCH	EWMA	Historical Mean	Implied Vol.	50 Day MA	200 Day MA	Best
MSE	1	24	5	4	13	4	EWMA
MAPE	0	19	6	4	18	4	EWMA
RMSE	1	24	5	4	13	4	EWMA
Abs of MFE	1	17	5	4	20	4	50 Day MA
LLF	0	19	5	4	19	4	EWMA/50DayMA

The table shows that both EWMA and the 50 Day Moving Average faired substantially better than the other models across all loss functions. EWMA was best across three of the loss functions, while the 50 Day Moving Average was best for Abs MFE, with EWMA and the 50 Day Moving Average tying for the number of times they had the lowest LLF value. Surprisingly, GARCH (1, 1) performed the worst out of all models across all loss functions.

Table 2 below shows the raw values for each of the loss functions for the aggregated series.

Table 2: Loss Function Values for Aggregated Non-Rolling Series

	GARCH	EWMA	Historical Mean	Implied Vol.	50 Day MA	200 Day MA	Best
MSE	0.00994%	0.00852%	0.00981%	0.01243%	0.00830%	0.00961%	50 Day MA
MAPE	1271.55	1096.98	1277.05	1525.32	1091.51	1291.78	50 Day MA
RMSE	0.9971%	0.9230%	0.9902%	1.1150%	0.9112%	0.9804%	50 Day MA
Abs of MFE	0.5443%	0.3670%	0.5374%	0.6550%	0.3485%	0.5068%	50 Day MA
LLF	3.0934	2.7065	3.0844	3.1911	2.6701	2.9795	50 Day MA

Table 2 shows that for the aggregated non-rolling series, the 50 Day Moving Average had the lowest values across all loss functions. It is safe to conclude that the 50 Day Moving Average was best at forecasting volatility from t=0 through t=23 on average, with all forecasts made at t=0 for each of the models except the two Moving Average models. As previously noted, the 50 and 200 Day Moving Averages *do roll* after each t (day) through each of the frequencies (otherwise they would not be moving averages). It is also apparent that the non-rolling EWMA performed only slightly worse across all loss functions than did the 50 Day Moving Average, and was thus clearly the second-best model for forecasting volatility on a non-rolling basis at t=0. The closeness in loss function values between EWMA and the 50 Day Moving Average is likely due to the similarity in the way in which the values for those models are calculated, as both are skewed towards favoring more recent data.

Weekly-Rolling Forecast

Table 3 below shows the number of times a model had the lowest value for a given loss function out of the total 10 assets for the weekly-rolling forecast.

Table 3: Number of Times Lowest Loss Function Value for Individual Assets (Weekly-Rolling)

	GARCH	EWMA	Mean	Implied Vol.	50 Day MA	200 Day MA	Best
MSE	0	4	2	1	2	1	EWMA
MAPE	1	4	1	1	2	1	EWMA
RMSE	0	4	2	1	2	1	EWMA
Abs of MFE	0	4	2	1	2	1	EWMA
LLF	0	4	2	1	2	1	EWMA

The table shows that EWMA had the lowest loss function four times for each loss function, with the long-run Mean and the 50 Day Moving Average tying for second place in all loss functions except MAPE. Again, GARCH (1, 1) performed the worst across the models. We are working with an admittedly low sample size for this rolling frequency due to computational limitations, which is why we aggregate the data into an individual series and evaluate the raw values of the loss functions for the aggregated series below.

Table 4 below shows the raw values for the loss functions for the aggregated series for the weekly-rolling forecasts.

Table 4: Loss Function Values for Aggregated Weekly-Rolling Series

	GARCH	EWMA	Mean	Implied Vol.	50 Day MA	200 Day MA	Best
MSE	0.00650%	0.00593%	0.00720%	0.00804%	0.00553%	0.00677%	50 Day MA
MAPE	1139.54	1054.57	1284.82	868.29	1072.10	1286.62	Implied Vol.
RMSE	0.8060%	0.7700%	0.8483%	0.8969%	0.7439%	0.8231%	50 Day MA
Abs of MFE	0.4003%	0.2856%	0.4773%	0.3980%	0.2766%	0.4325%	50 Day MA
LLF	2.8772	2.5873	3.0757	2.5826	2.5937	2.9436	Implied Vol.

The table shows that the 50 Day Moving Average had the lowest values for three loss functions (MSE, RMSE, and Abs MFE) and that Implied Volatility had the lowest value for two (MAPE and LLF). Notably, for the loss functions for which the 50 Day Moving Average had the lowest value, EWMA was again the second-best performer, performing only slightly worse than the 50 Day Moving Average; this is again likely due to the similarity in how these models are calculated.

Daily-Rolling Forecast

Table 5 below shows the number of times a model had the lowest value for a given loss function out of the total 10 assets for the weekly-rolling forecast.

Table 5: Number of Times Lowest Loss Function Value for Individual Assets (Daily-Rolling)

	GARCH	EWMA	Mean	Implied Vol.	50 Day MA	200 Day MA	Best
MSE	1	0	1	2	5	1	50 Day MA
MAPE	0	9	0	0	1	0	EWMA
RMSE	1	0	1	2	5	1	50 Day MA
Abs of MFE	3	0	0	2	4	1	50 Day MA
LLF	2	0	1	1	6	0	50 Day MA

Again, as we are working with a small sample size of only 10 assets, these rankings should be taken with a grain of salt. However, the table does demonstrate that the 50 Day Moving Average had the lowest value for each loss function the most times for every loss function except MAPE. GARCH (1, 1) improved noticeably relative to other forecast frequencies when producing 1-step ahead forecasts.

Table 6 below shows the raw values for the loss functions for the aggregated series for the daily-rolling forecasts.

Table 6: Loss Function Values for Aggregated Daily-Rolling Series

	GARCH	EWMA	Mean	Implied Vol.	50 Day MA	200 Day MA	Best
MSE	0.00480%	0.01027%	0.00577%	0.00684%	0.00471%	0.00489%	50 Day MA
MAPE	56194.97	186.01	78706.51	312.03	52129.04	75069.76	EWMA
RMSE	0.6927%	1.0134%	0.7596%	0.8269%	0.6862%	0.6996%	50 Day MA
Abs of MFE	0.2404%	0.6706%	0.3602%	0.2546%	0.1125%	0.3001%	50 Day MA
LLF	3.5360	79.2766	3.8376	1.6756	3.3497	3.6787	Implied Vol.

The table shows that again the 50 Day Moving Average had the lowest loss function values for three of the loss functions, while both EWMA and Implied Volatility had the lowest value for one loss function each. GARCH (1, 1) had the second lowest values for MSE, RMSE, and Abs MFE, again a notable improvement from its performance on longer-horizon forecasts.

Table 7 presents the DM test statistics for the linear loss functions (MAPE and Abs MFE), with the statistically significant statistics boldfaced (recall a DM test statistic is significant if it is greater in absolute value than 1.96).

Table 7: Diebold Mariano Test Statistics Matrix for Linear Loss Functions

	GARCH	**EWMA**	**Mean**	**Implied Vol.**	**50 Day MA**	**200 Day MA**
GARCH		-1.6276	**-4.0033**	**-2.4845**	0.4189	-1.0083
EWMA	-1.6276		0.6599	1.2241	**2.0324**	1.2847
Mean	**-4.0033**	0.6599		**2.3489**	**2.7274**	**2.9030**
Implied Vol.	**-2.4845**	1.2241	**2.3489**		1.7570	**2.0065**
50 Day MA	0.4189	**2.0324**	**2.7274**	1.7570		-1.0706
200 Day MA	-1.0083	1.2847	**2.9030**	**2.0065**	-1.0706	

The DM test statistics of the linear loss functions are clearly inconclusive. The table does not show that any single model has a consistently statistically significant difference in linear loss function values. Importantly, the 50 Day Moving Average has significantly different linear loss function values only between itself and EWMA and the long-term Mean.

Table 8 presents the DM test statistics of the quadratic loss functions (MSE, RMSE, and LLF), with the statistically significant statistics again boldfaced.

Table 8: Diebold Mariano Test Statistics Matrix for Quadratic Loss Functions

	GARCH	**EWMA**	**Mean**	**Implied Vol.**	**50 Day MA**	**200 Day MA**
GARCH		**-2.0563**	**-2.3527**	-0.9189	0.2187	-0.1614
EWMA	**-2.0563**		1.5592	**2.1743**	**2.2923**	1.7400
Mean	**-2.3527**	1.5592		1.5169	1.6825	**2.1339**
Implied Vol.	-0.9189	**2.1743**	1.5169		0.4740	1.1130
50 Day MA	0.2187	**2.2923**	1.6825	0.4740		-0.2390
200 Day MA	-0.1614	1.7400	**2.1339**	1.1130	-0.2390	

The DM test statistics of the quadratic loss functions are also inconclusive. EWMA, however, does have statistically significantly different loss function values amongst itself and GARCH, Implied Volatility, and the 50 Day Moving Average (the 200 Day Moving Average would be accepted with a lower critical value).

Market Context and Analysis

US equities are currently in the second longest bull market in history, spanning a period of over eight years. Since the election of President Donald Trump on November 8, 2017, developed-economy market indices around the world have rallied, with the S&P 500 climbing just shy of 10% from the election to March 8, 2017. The index also traded in a sub-1% range for 50 consecutive days, one of the least volatile streaks in its history. Even more relevant to the purposes and results of this paper, the volatility index (VIX) is at a decade low, as shown below in Figure 1.

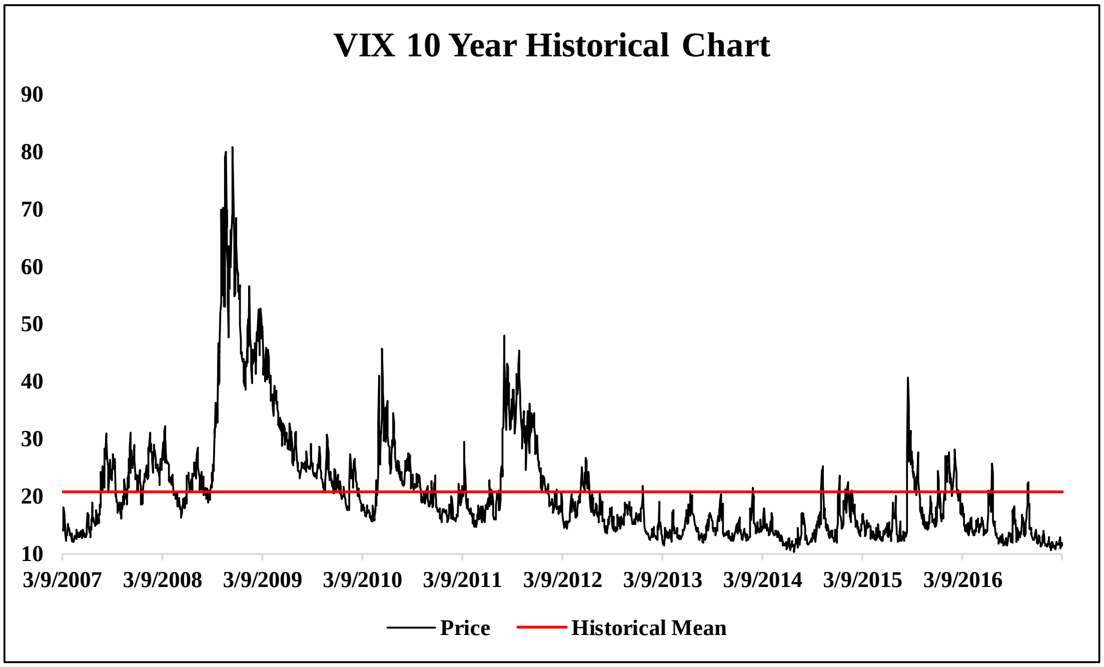

Figure 1: VIX 10 Year Historical Chart

Figure 2 below zooms in on Figure 1 to highlight the volatility of the overall market over the last year (up through 3/9/2017).

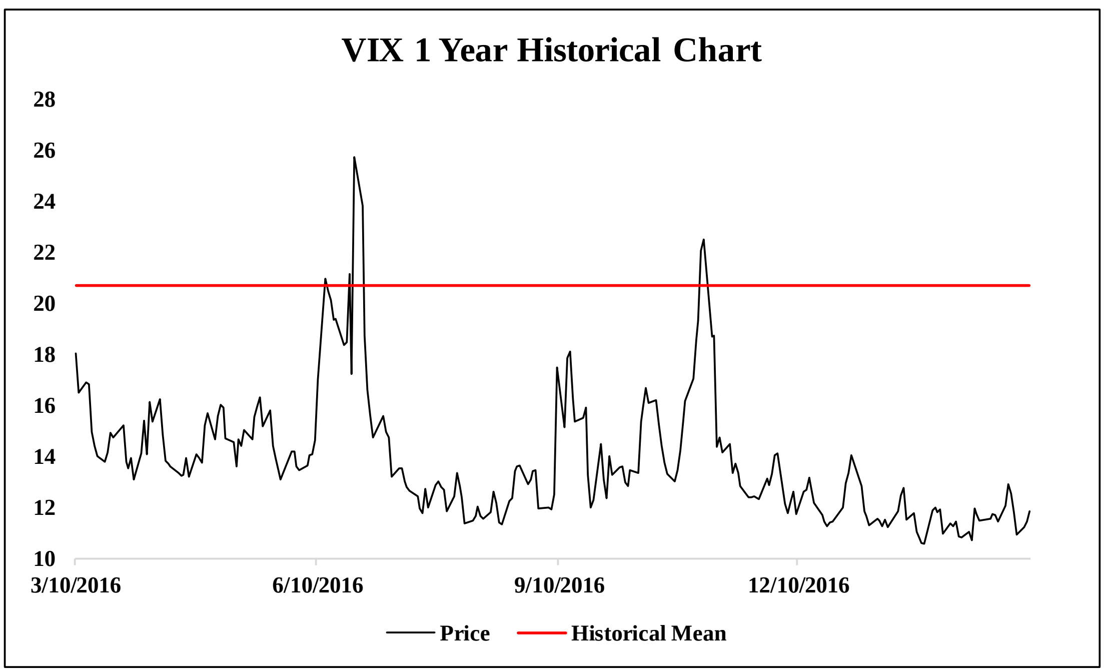

Figure 2: VIX 1 Year Historical Chart

Figure 2 highlights the fact that the VIX was at a decade low during the month of February, our forecast period. In other words, our forecast period was a period of abnormally low volatility and all-time stock market highs. With this in mind, the results of the study and the values of the loss functions are not very surprising. Specifically, the 50 Day Moving Average was likely the best performer because it incorporated *only* data from this period of abnormally low volatility, and similarly, EWMA was also a good performer because it gives more weight to the most recent observations of volatility. Unlike the 50 Day Moving Average, the 200 Day Moving Average considers data pre "Trump Rally," which clearly diminished the model's efficacy as an accurate predictor of volatility for the forecast period. This also offers a potential explanation as to why GARCH (1, 1) performed much more poorly than expected: one of the largest assumptions of GARCH is reversion to a long-term mean level of volatility, which did not happen at all, even for individual assets, in this forecast period.

Given the context of the broader market and the lack of clear statistical significance in the results, it is best to say that we can still not conclude which of the chosen models is the best forecaster of volatility. However, that does not mean that this study was unfruitful. We can conclude that 50 Day Moving Averages and EWMA volatility measures produce the *better* forecasts as measured by loss function values across differing time horizons than the other models in the study during periods of abnormally low volatility. We can also conclude that GARCH (1, 1) is relatively ineffective at predicting volatility during periods of low volatility for longer forecast horizons.

Interestingly, the results of this study were abnormal relative to the empirical studies reviewed in Chapter 2. Many of these studies produced results where GARCH or a GARCH-family model were the most accurate predictor of future volatility. However, it is important to note that few empirical studies incorporated moving averages in their tests. This could easily lead one to conclude that the criticism presented in Chapter 2 that mathematically complex models offer little value relative to simpler models is validated by the results of this study. Again, however, one should bear in mind the broader market context and ex-post realization of abnormally low volatility and one should also recognize the differences in methodology and forecast window/rolling frequency among this paper and others. The results did show a slight improvement for GARCH (1, 1) with the daily-rolling frequency, suggesting that GARCH performs better on shorter time horizons (although it still was not the best forecaster). Other empirical studies which have tested models on frequencies such as intra-day have shown that GARCH is the most accurate predictor of volatility on such frequencies. Although this study did not look at intra-day data, due to the overwhelmingly supportive empirical data from other studies such as those reviewed in Chapter 2, practitioners looking to forecast volatility on intra-day data should use GARCH (1, 1).

Chapter 6

Summary and Conclusion

In this paper, we reviewed a sampling of the vast literature on the subject of volatility forecasting. We then selected models which are accessible to the average trader and portfolio manager to test over various forecast horizons and rolling methodologies. We tested these models on a sample of assets from various asset classes. We then ranked the performance of these models via commonly used statistical loss functions and tested the statistical significance of the difference in loss functions between models for the daily-rolling forecast frequency.

We conclude that the results are relatively inconclusive and do not inform us as to which model is the best at predicting volatility due to the lack of clear statistical significance and outperformance amongst the models. We consider the broader market dynamics for the month of February, the chosen forecast period, and note that the broader market was in a period of historically low volatility during the period of study. We reason that given the market context allows us to conclude that models which favor recent observations of volatility are *better* predictors of future volatility than models such as GARCH (1, 1) which have an underlying assumption of mean reversion during periods of abnormally low volatility. Thus, the results suggest that practitioners use simple moving averages such as the 50 Day Moving Average volatility during prolonged periods of abnormally low volatility if they believe markets are likely to stay generally calm for their forecast period.

Further research should seek to replicate and improve upon the methodology of this paper. Namely, future studies should forecast in real-time as opposed to out of sample and should use different rolling frequencies to determine the efficacy of models across various forecast horizons. More specifically, other studies should seek to extend the methodology of this paper over a longer time horizon (such as for an in-sample period of one year as opposed to one month as was the case in this paper) and over various levels of general market volatility. In other words, future studies should test across multiple estimation windows and various market environments in order to produce more reliable results. Future studies, while also including the models tested in this paper, should seek to include other volatility models which were not included in this study, such as Autoregressive Moving Averages (ARMA), E-GARCH, or other n-day moving averages which may be better forecasters of volatility than those chosen for this study. Other studies should attempt to perform Diebold Mariano tests of statistical significance on loss functions between models for all rolling frequencies, assuming one can successfully implement a multivariate DM test which accounts for all error terms.

BIBLIOGRAPHY

Andersen, T. G., & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review,39*(4), 885. doi:10.2307/2527343

Berestycki, H., Busca, & Florent, I. (2004). Computing the implied volatility in stochastic volatility models. *Communications on Pure and Applied Mathematics,57*(10), 1352-1373. doi:10.1002/cpa.20039

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, *81*(3), 637-654. doi:10.1086/260062

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307-327. doi:10.1016/0304-4076(86)90063-1

Day, T. E., & Lewis, C. M. (1993). Forecasting Futures Market Volatility. *The Journal of Derivatives*, *1*(2), 33-50. doi:10.3905/jod.1993.407876

Efimova, O., & Serletis, A. (2014). Energy markets volatility modelling using GARCH. *Energy Economics, 43,* 264-273. doi:10.1016/j.eneco.2014.02.018

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987. doi:10.2307/1912773

Figlewski, S. (1997). Forecasting Volatility. *Financial Markets, Institutions and Instruments*, *6*(1), 1-88. doi:10.1111/1468-0416.00009

Forde, M., Jacquier, A., & Mijatovic, A. (2010). Asymptotic formulae for implied volatility in the Heston model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 466*(2124), 3593-3620. doi:10.1098/rspa.2009.0610

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics, 20*(7), 873-889. doi:10.1002/jae.800

Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies,6*(2), 327-343. doi:10.1093/rfs/6.2.327

Hull, J., & White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance, 42*(2), 281. doi:10.2307/2328253

Hull, J. (2009). *Options, futures, and other derivatives*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Khalifa, A. A., Miao, H., & Ramchander, S. (2010). Return distributions and volatility forecasting in metal futures markets: Evidence from gold, silver, and copper. *Journal of Futures Markets, 31*(1), 55-80. doi:10.1002/fut.20459

Malliavin, P., & Mancino, M. E. (2009). A Fourier transform method for nonparametric estimation of multivariate volatility. *The Annals of Statistics,37*(4), 1983-2010. doi:10.1214/08-aos633

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business, 36*(4), 394. doi:10.1086/294632

Marra, S. (2015). *Predicting Volatility* (Rep.). Lazard Asset Management.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347. doi:10.2307/2938260

Pagan, A. R., & Schwert, G. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*(1-2), 267-290. doi:10.1016/0304-4076(90)90101-x

RiskMetrics technical document. (1995). New York: JP Morgan.

Zhang, L., Mykland, P., & Ait-Sahalia, Y. (2003). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data. doi:10.3386/w10111

ACADEMIC VITA
CHRISTIAN N. BINDER

CNB5189PSU@GMAIL.COM

EDUCATION

The Pennsylvania State University | Schreyer Honors College **University Park, PA**
Smeal College of Business **|** Bachelors of Science in Finance *Class of 2017*
College of the Liberal Arts **|** Bachelors of Science in Economics, History Minor

Maastricht University **Maastricht, Netherlands**
Center for European Studies *Jun 2014 – Aug 2014*

RELEVANT EXPERIENCE

Bank of America Merrill Lynch **New York, New York**
Investment Banking Summer Analyst, Mergers & Acquisitions *Jun 2016 – Aug 2016*

- Worked on five deals including sell-side and buy-side M&A across multiple sectors such as Healthcare and Consumer & Retail
- Aided analyst in the construction of the financial model and the preparation of presentation materials for the announced $624 MM sale of Cynapsus Therapeutics to Sunovion Pharmaceuticals which were presented to the client's management and Board of Directors
- Assisted with a clinical laboratory's potential $4 bn acquisition of a contract research organization
- Researched potential UK acquisition and investment opportunities based on post-Brexit performance relative to EU and UK revenue exposure which was included as part of a newsletter sent to all employees at Bank of America Merrill Lynch

Wells Fargo Securities, LLC **Charlotte, NC**
Investment Banking Summer Analyst, Investment Grade Loan Syndications *Jun 2015 – Aug 2015*

- Evaluated the investment grade loan market by analyzing current pricing trends and the volume of new deals vs. refinancings by using information from online databases to compose graphs and charts which were presented to the group and prospective clients
- Worked on a deal team for a $1.5 billion refinancing of a large independent oil and gas company's revolver
- Completed a case study involving an analysis of a potential acquisition by a biotech company of a middle-market contract research organization, which included a DCF, merger model, and precedent transactions analysis which was presented to senior bankers

Nittany Lion Fund, LLC **University Park, PA**
Director of NLF Education *Jan 2016 – Present*

- Developed a "real-world application" focused formal education program to train over 40 Nittany Lion Fund Managers in five core lessons on Excel, Accounting, Financial Statement Projections, DCF Modeling, and Merger Modeling

Nittany Lion Fund, LLC **University Park, PA**
Lead and Associate Fund Manager, Consumer Discretionary Sector *Nov 2014 – Jan 2016*

- Co-managed the $960,000 Consumer Discretionary Sector within the $7.0 MM Nittany Lion Fund by using DCF models, ratios, and comparables to analyze seven holdings with the goal of creating a portfolio that outperforms the S&P 500
- Composed over 20 stock pitch presentations using fundamental analysis and long-term drivers to explain investment theses for buy, sell and hold recommendations to the Executive Board, Fund Managers, and the Penn State Investment Association

Penn State Investment Association **University Park, PA**
Analyst, Consumer Discretionary Sector *Aug 2014 – Present*

- Participated in education sessions to receive training in valuation, DCFs, comparables, ratios, stock pitches and research reports
- Valued, researched and pitched a retail company with a partner for a stock pitch competition against 30 other teams

Wall Street Boot Camp **University Park, PA**
Graduate *Aug 2014 – Dec 2014*

- Accepted into an exclusive group of 40 students from among hundreds of applicants to participate in weekly sessions presented by Wall Street professionals and Penn State alumni to prepare for a career in the financial services industry

LEADERSHIP EXPERIENCE

Student Government Association **Abington, PA**
Treasurer *Apr 2013 – Apr 2014*

- Recruited over 100 new members through rigorous marketing campaign, resulting in the largest membership in club history
- Chaired Abington's $40,000 Budget Committee, serving in a devil's advocate role to fund over 50 club events

Alpha Kappa Lambda Fraternity, PA Eta Colony **Abington, PA**
Recruitment Director *Aug 2012 – Apr 2014*

- Co-founded the first and only fraternity on the Abington campus with nine other brothers; member of the first initiation class
- Wrote an intensive business plan to quadruple membership over a two year period with the goal of chartering into a chapter

ADDITIONAL INFORMATION

Honors: Beta Gamma Sigma, Dean's List, PSU Summer Leadership Conference, Abington President's Leadership Retreat
Computer Skills: R, VBA, STATA, Bloomberg Terminal, FactSet, CapitalIQ, @RISK