10-20-2017

Evaluating Volatility Forecasts in Various Equity Market Regimes

John P. Felletter

DISSERTATION
Number DBA06/2017

Evaluating Volatility Forecasts in Various Equity Market Regimes

Submitted by

John P. Felletter

Doctor of Business Administration in Finance Program

In partial fulfillment of the requirements

For the degree of Doctor of Business Administration in Finance

Sacred Heart University, Jack Welch College of Business

Fairfield, Connecticut

Date: October 20, 2017

Dissertation Supervisor: Dr. W. Keener Hughen Signature: W. Keener Hughen

Committee Member: Dr. Lucjan T. Orlowski Signature:

Committee Member: Dr. I-Hsuan Ethan Chiang Signature:

Evaluating Volatility Forecasts in Various Equity Market Regimes

John P. Felletter, CFA, FRM
Doctor of Business Administration in Finance Dissertation
John Welch College of Business
Sacred Heart University

ABSTRACT

Forecasting volatility is a critical component of asset allocation, risk management, and option pricing. Many different methods and models are used to predict volatility, and many studies have examined the efficacy of one method or another. This study investigates whether the abilities of historical volatility, GARCH models, and VIX to forecast volatility vary in different market conditions, as distinguished by levels of volatility and returns. It is found that market conditions do impact the abilities of the variables to forecast volatility. Overall, the forecasts implied by the GARCH models perform best according to the various metrics, while the VIX forecast has the worst performance. However, with a simple linear correction for the bias, the forecasts implied by VIX and the leverage GARCH model perform best. Their relative performance depends on market conditions; the VIX forecast performs best during times when volatility is low. These results indicate that practitioners who wish to forecast volatility should take current market conditions into account.

JEL Classification C22, C52, E37, G12

I. INTRODUCTION

Forecasting volatility is a crucial component of portfolio management, risk management, asset pricing, and option trading. Multiple types of models have been used to predict volatility. Approaches include historical volatility, the CBOE S&P 500 VIX model (hereinafter "VIX"), ARCH/GARCH class conditional volatility models, and stochastic volatility models such as the Heston model. This study evaluates a historical model, two GARCH models: GARCH (1,1) and the leverage GARCH(1,1), and VIX. The forecasts provided by these models are compared to actual realized returns of the S&P 500 to determine if the models perform differently, both absolutely and relative to each other, in different market conditions, or "regimes."

The field of finance is fundamentally about risk and uncertainty, and volatility is the most commonly used measure of uncertainty. As Campbell, Lo, and MacKinley (1997) state, "The starting point for every financial model is the uncertainty facing investors, and the substance of every financial model involves the impact of uncertainty on the behavior of investors and, ultimately, on market prices. The very existence of financial economics as a discipline is predicated on uncertainty."

Since volatility is a common proxy for risk, an ability to predict volatility is a, if not the, critical ability in the development of risk management models. Many risks, liquidity risk, inflation risk, exchange risk, credit, market risk etc., can impact portfolio performance. However, if it was known with certainty the direction that of any risk, it would be relatively easy to hedge against any risk that was of concern to a particular portfolio. The uncertainty of the movement of these factors, which could be described as the volatility in these factors, eliminates the ability to directly hedge. Historical volatility is the key input into VAR models, but an improvement in the predictability of volatility would result in a direct improvement of risk modeling.

The Sharpe-Lintner Capital Asset Pricing Model ("CAPM") (Sharpe, 1964, Lintner 1965) was predicated on the assumption that all investors should hold mean-variance efficient portfolios, where mean is the expected return, and variance is the variance of returns within the portfolio. Investors try to obtain the maximum return for a given level of risk, or alternatively, accept a minimum level of risk for a given return. Fama and French (1992) expanded upon this by adding size and value factors, but the risk/return trade-off remained the primary component of asset pricing. These approaches that attempt optimal portfolio construction by maximizing risk/return throughout the spectrum are based upon one fundamental belief, that the risk in any given investment can in fact be assessed.

There are a number of ways to trade volatility. The simplest way would be to go long or short, as appropriate, a traded volatility proxy, such as the VIX. Option sellers are short volatility, as they are betting that the underlying stays within a ban inside of the strike prices. Option buyers, alternatively, are long volatility, hoping that the underlying pricing breaks out beyond the strike price. More sophisticated derivatives such as variance and volatility swaps are also used both as hedges, and as ways to trade based on a belief in the future level of volatility. Clearly it is axiomatic that someone trading volatility would want to be able to predict volatility, to understand if the particular traded instrument is appropriately priced, and ideally to exploit a pricing inefficiency.

Numerous variables and methods have been used to forecast market volatility, as measured by realized volatility of the S&P 500 index. These variables are constructed from past volatility or are forecasts implied from option prices. In particular, the VIX, which is the model-free implied volatility constructed from nearby options on the S&P 500 index, is often used as a forecast of market volatility, as is past realized volatility, i.e., historical volatility. Both VIX and historical volatility are likely to be biased forecasts of realized volatility. The VIX is constructed from option prices and hence is the risk-adjusted expectation of realized volatility; if investors view volatility risk either favorably or unfavorably then the VIX will differ from the actual expectation because of the nonzero risk premium. Historical volatility is past realized volatility. If volatility followed a random walk, then historical volatility would be an unbiased forecast; however, volatility has been shown to display mean reversion and in particular does not appear to follow a random walk.

Generalized autoregressive conditional heteroskedastic, or GARCH, models are a relatively parsimonious class of models that allow for many of the stylized features of volatility. Importantly, GARCH models give formulas for expected realized volatility and therefore their forecasts of volatility can be computed in closed form. Their forecasts can thus be compared with the more naïve VIX and historical volatility forecasts.

While numerous papers have attempted to evaluate which method is the best for predicting volatility, to our knowledge what has not been done in the literature is to compare the forecasts in different market conditions. This paper investigates two questions, whether the success of these variables for predicting volatility is a function of underlying market conditions, as distinguished by levels of returns and volatility, and whether different methods of predicting volatility perform relatively better in these different conditions.

It is found that, as expected, neither VIX nor historical volatility are accurate forecasts of realized volatility. GARCH-based models, specifically standard GARCH (1,1), and Leveraged GARCH (1,1), consistently outperformed the VIX model when evaluated using typical forecast evaluation tools, and the Diebold-Mariano model.

This remainder of this paper is presented as follows: Section II addresses prior research, Section III addresses data selection and methodology, Section IV contains findings, and Section V contains the conclusion, and recommendations for further research.

II. Literature Review

Poon and Granger (June 2003), prepared a comprehensive review of 93 papers that dealt with volatility forecasting written over the previous two decades (from 1983 through 2003). Note that rather than presenting a new approach to volatility forecasting, they focus on summarizing research to date, and, from the universe of existing research, attempt to draw conclusions. They look at two overall questions, is volatility forecastable; and, which approach is best. They state that "Volatility is not the same as risk….(but) A good forecast of the volatility of asset prices over the investment holding period is a good starting point for assessing investment risk." Further they assert the importance of predicting volatility for both option pricing and portfolio risk management. They start by defining, and comparing and contrasting, key terms, volatility, standard deviation, and risk. Then give brief overviews of the mechanics of Time Series (Historical) models, ARCH/GARCH type models Stochastic Models, and Option-Implied Models. Similar to this paper, they compare the results of historical volatility approaches to those using volatility implied from options. In a summary paper published in the Financial Analysts Journal, Volume 61, Number 1, 2005, they conclude that "*Based on the forecasting results, option implied volatility* *dominates time-series models because the market option price fully incorporates current information and future volatility expectations. Between historical volatility and ARCH models, we found no clear winner, but they are both better than the stochastic volatility model. Despite the added flexibility and complexity of SV models, we found no clear evidence that they provide superior volatility forecasts. Also, high frequency data clearly provide more information and produce better volatility forecasts, particularly over short horizons."*

Brownlees, Engle, and Kelly, (2011) evaluated predictive models, within the ARCH class of models, to *"identify successful predictive models over multiple horizons and to investigate how predictive ability is influenced by choices for estimation window length, innovation distribution, and frequency of parameter re-estimation."* As mentioned, in their analysis they used five "ARCH" based models, GARCH (1,1), the basic GARCH model, TARCH, Threshold ARCH, which addresses the tendency for volatilities to increase when past returns are negative by appending a linear asymmetry adjustment, EGARCH, Exponential GARCH, which models the low of the variance, NGARCH, Nonlinear GARCH, addresses negative returns by amplifying negative news relative to positive news, and, APARCH, Asymmetric power ARCH. Brownlees, et al found that, asymmetric models perform well across all methods, assets, and sub-samples. And that models perform best using the longest available data series; and, updating parameters regularly helps counteract the adverse effects of parameter drift.

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H., (June 2014) evaluated the findings or recent research that indicates that the variance risk premium predicts aggregate stock market returns. Their findings corroborated the view that such predictability exists. Further, they identified a "global" variance risk premium, and found that a similar pattern of predictability existed, in markets such as France, Germany, Japan, Switzerland, the Netherlands, Belgium, and the United Kingdom.

Bekaert and Hoerova (2014) attempted to decompose the squared VIX index into two components, the conditional variance of stock returns, and the equity variance premium (the difference between the squared VIX index, and an estimate of the conditional variance of the stock market). They describe the VIX as follows: "The VIX index is the "risk-neutral" expected stock market variance for the US S&P500 contract and is computed from a panel of options prices". They state that VIX, also known as the "fear index" (Whaley, 2000), reflects both stock market uncertainty, and a variance risk premium. Measuring the variance risk premium is difficult however, because it requires an estimate of the conditional variance of stock returns. They concluded that while the variance premium was a "significant predictor of stock returns", the conditional variance was not.

Brownlees and Gallo, (2010) using Ultra High Frequency Data ("UHFD") apply different historical volatility measures, realized volatility, bi-power realized volatility, two-scales realized volatility, realized kernel, and the daily range, to forecast Value at Risk (VaR). They found that: "*Simple benchmarks such as GARCH, unconditional variance, historical simulation, and RiskMetrics are clearly outperformed when UHFD volatility measures and the range are used, confirming the point that absolute or squared returns are less efficient as estimates of the relevant variance. We find that UHFD volatility measures and the daily range perform similarly well in terms of VaR forecasting, with the former obtaining the best forecasting results at "low" frequencies (20 or 30 minutes)."*

Todorov, (2010) examined the variation over time in the market variance-risk premium. He concluded that "jumps play a key role in explaining the variance risk premium". While port jump spikes in volatility "die out quickly", there is a persistent increase in the variance risk premium. This finding could imply that, as it relates to option pricing, a time series predictor could provide an accurate estimate if the model correctly evaluates the impact of recent jumps.

III. DATA SELECTION AND METHODOLOGY

The data consist of daily returns on the S&P 500 Index (SPX), computed from daily closing values obtained from S&P Capital IQ, and daily observations of the VIX index, obtained from the Chicago Board Options Exchange (CBOE) website. The sample period is January 2, 1990, when the VIX data begin, to December 31, 2016.

The volatility to be forecast is the 22-day realized volatility:

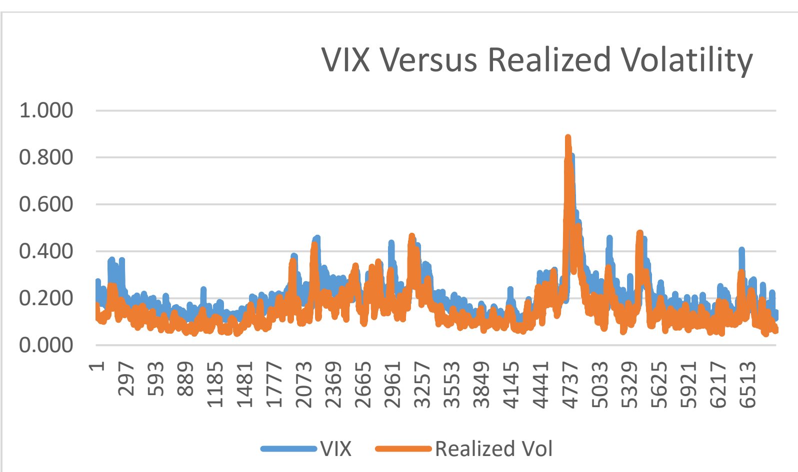
$$RV_{t,22} = \sqrt{\frac{1}{22}\sum_{n=1}^{22}(R_{t+n} - \bar{R}_t)^2} \quad (1)$$

where R_s is the S&P 500 return over day s and $\bar{R}_t$ is the average daily S&P 500 return over the 22-day window. All returns are annualized. Note that today's realized volatility, $RV_{t,22}$, uses data over subsequent days and is thus not actually observed until 21 days later. For example, the realized volatility on January 1 is not observed until January 21.

Table 1 - Realized Volatility

REALIZED VOLATILITY – KEY STATISTICS	
Total Inputs	6805
Average Daily Volatility - Annualized	15.46%
Maximum Daily Volatility - Annualized	88.68%
Minimum Daily Volatility - Annualized	4.64%
Standard Deviation of Daily Volatility	9.15%
Percentage of Dates in which VIX Prediction > Realized Volatility	85.8%
Autocorrelation	0.994

Figure 1



This measure of volatility is used for several reasons. First, it is the measure of volatility that is typically forecast in related studies[1]. Second, it can be interpreted as a proxy for the square root of the average variance over the subsequent month, which is typically used as the underlying variable in a variance swap contract. Third, the model free implied variance, computed from option prices, is closely related to this measure as explained in more detail below and can be interpreted as the measure's risk-adjusted forecast.

If realized volatility were a martingale, for example, if it followed a random walk, then the best forecast of future realized volatility would be past realized volatility. As shown in Table 1, *RV* is highly persistent and looks to be very nearly a random walk. Thus a natural candidate to forecast today's realized volatility is the realized volatility that is most recently observed, today's *historical volatility*, computed using return data over the previous 22 days:

$$HV_{t,22} = \sqrt{\frac{1}{22}\sum_{n=0}^{21}(R_{t-n} - \bar{R}_t)^2} \tag{2}$$

Despite the high persistence of realized volatility, theoretical considerations and previous research provide strong evidence that volatility tends to mean-revert and that returns are heteroskedastic. If returns exhibit heteroskedasticity then an equally weighted average of squared deviations from the mean, as defined in Equation (2), may not be optimal. To test the extent of heteroskedasticity within the data, the Breusch-Pagan-Godfrey test, the Harvey test, and the Glejser test were run. The results are given in Table 2:

Table 2. Heteroskedasticity Tests

Breusch-Pagan-Godfrey

F-statistic	368.9582	Prob. F(1,6803)	0.0000
Obs*R-squared	350.0802	Prob. Chi-Square(1)	0.0000
Scaled explained SS	3651.733	Prob. Chi-Square(1)	0.0000

Harvey

F-statistic	685.1546	Prob. F(1,6803)	0.0000
Obs*R-squared	622.6470	Prob. Chi-Square(1)	0.0000
Scaled explained SS	709.0051	Prob. Chi-Square(1)	0.0000

[1] See, for example, Bekaert and Hoerova (2014), Chernov (2001), and Poon and Granger (2003).

Glejser

F-statistic	997.1826	Prob. F(1,6803)	0.0000
Obs*R-squared	869.9575	Prob. Chi-Square(1)	0.0000
Scaled explained SS	1452.780	Prob. Chi-Square(1)	0.0000

All three tests strongly reject the null hypothesis of homoscedasticity.

Generalized autoregressive conditional heteroskedasticity (GARCH) models allow for heteroskedasticity, as well as volatility clustering and long-run mean-reversion. These models explicitly model the variance of daily return innovations as a separate variable whose evolution depends on its past values and on past return innovations. In particular, in the GARCH(p,q) model the conditional variance $\sigma_t^2 = var_{t-1}(R_t)$ evolves according to:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \, \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \, \sigma_{t-i}^2 \qquad (3)$$

where ε_t is the return innovation:

$$R_t = \mu + \varepsilon_t \qquad (4)$$

To determine the optimal number of lags p and q, the Akaike Information Criterion (AIC) and Schwartz Information Criterion (SIC) were calculated for various lags. The results are shown in Table 3.

Table 3.

Model	Akaike Info Criterion ("AIC")	Schwarz Criterion ("SIC")
GARCH (1,1)	**-6.54**	**-6.54**
GARCH (0,1)	-6.24	-6.23
GARCH (0,2)	-6.37	-6.36
GARCH (0,3)	-6.42	-6.41
GARCH (0,9)	-6.53	-6.51
GARCH (2,1)	-6.54	-6.54
GARCH (3,1)	-6.54	-6.54
GARCH (4,1)	-6.54	-6.53
GARCH (5,1)	-6.54	-6.53

The AIC and SIC both indicate that the more parsimonious GARCH(1,1) model is appropriate:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad (5)$$

The variance σ_t^2 in GARCH models is the variance of the return over day t. The average daily variance over the subsequent 22 days is approximately equal to the realized variance (the square of the realized volatility):

$$RV_{t,22}^2 \approx \frac{1}{22}\sum_{n=0}^{21}\sigma_{t+n+1}^2 \tag{6}$$

Therefore the GARCH forecast of realized variance is the sum of expected future variances. Fortunately, GARCH models have the very useful property that expected future variance can be written as a linear function of current variance. For example, the expected one-step ahead variance is

$$E_t[\sigma_{t+2}^2] = \omega + \alpha E_t[\varepsilon_{t+1}^2] + \beta E_t[\sigma_{t+1}^2] = \omega + (\alpha+\beta)\sigma_{t+1}^2 \tag{7}$$

Similarly, the formula for the expected n-step ahead variance is

$$E_t[\sigma_{t+n+1}^2] = \omega\frac{1-\gamma^n}{1-\gamma} + \gamma^n\sigma_{t+1}^2, \text{ where } \gamma = \alpha+\beta. \tag{8}$$

See, for example, Hull (2015) for details. Making use of the formula for the *N*th partial sum of a geometric series, the sum of the expected variances for the subsequent *N* days can also be written as a linear function of current variance:

$$E_t\left[\sum_{n=0}^{N-1}\sigma_{t+n+1}^2\right] = \frac{\omega}{1-\gamma}\left(N - \frac{1-\gamma^N}{1-\gamma}\right) + \frac{1-\gamma^N}{1-\gamma}\sigma_{t+1}^2 \equiv A(N) + B(N)\sigma_{t+1}^2 \tag{9}$$

It follows that expected realized volatility is given by

$$E_t\big[RV_{t,22}\big] = \sqrt{\frac{1}{22}\left(A(22) + B(22)\sigma_{t+1}^2\right)} \tag{10}$$

Today's expectation of a random variable is the forecast of that variable; therefore the forecast of realized volatility according to the GARCH model defined in Equation (5) is the expectation in Equation (10).

The standard GARCH(1,1) model defined in Equation (5) has symmetric response to return innovations ε_{t-1}: next period's variance increases by the same value whether the innovation is positive or negative. However, much empirical evidence supports the so-called "leverage effect" that return innovations are negatively correlated with volatility. For example, with the data sample used in this study, the correlation between return innovations and realized volatility is -10.4% and the correlation between return innovations and VIX is -11.8%. To account for this effect, the leverage GARCH model is considered. This model is defined by

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2(\hat{\varepsilon}_{t-1} - \theta)^2 + \beta\sigma_{t-1}^2, \text{ where } \hat{\varepsilon}_{t-1} = \varepsilon_{t-1}/\sigma_{t-1} \tag{11}$$

For positive α and θ, a positive return innovation results in a smaller increase in variance next period than does a negative return innovation of the same magnitude, and thus volatility will tend to be greater when returns are negative.

The leverage GARCH model defined by Equation (11) also has the property that expected future variance is a linear function of current variance, with the same functional form as Equation (8) but with $\gamma = \alpha(1 +$

$\theta) + \beta$, and therefore the forecast of realized volatility for the leverage GARCH model takes the same form as Equations (9) and (10).

The forecasts of realized volatility provided by historical volatility and GARCH models as described above are based on past data. In contrast, volatility implied from option prices are forward looking and represent risk-adjusted expected future volatility. The Chicago Board Options Exchange Volatility index, *VIX*, is constructed from the highly liquid S&P 500 index options with roughly one month to maturity. Its squared value is the model free implied variance and thus is the risk-adjusted forecast of 22-day realized variance, $RV_{t,22}^2$.[2] It follows that the *VIX* provides a natural forecast of realized volatility.

This study evaluates and compares the abilities of these four variables to forecast realized volatility:

i) *HV* -- historical volatility
ii) *SGARCH* --the standard GARCH(1,1) model forecast defined in Equations (9) and (10) with $\gamma = \alpha + \beta$
iii) *LGARCH* --the leverage GARCH(1,1) model forecast also defined in Equations (9) and (10) but with $\gamma = \alpha(1 + \theta) + \beta$
iv) *VIX*

The forecasting performances of the variables are measured and compared using standard forecasting tests: root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), symmetric mean percentage error (SMPE), Theil U1, and a modified Theil U2 (explained below). The definitions of these metrics, as provided by the Department of Treasury(2008) are as follows. For each variable, let f_t denote its forecast of $RV_{t,22}$.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(RV_t - f_t)^2} \tag{12}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|RV_t - f_t| \tag{13}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|RV_t - f_t|}{RV_t} \tag{14}$$

$$SMAPE = \frac{2}{T}\sum_{t=1}^{T}\frac{|f_t - RV_t|}{f_t + RV_t} \tag{15}$$

$$U1 = \frac{\sqrt{\sum_{t=1}^{T}(RV_t - f_t)^2}}{\sqrt{\sum_{t=1}^{T}RV_t^2} + \sqrt{\sum_{t=1}^{T}f_t^2}} \tag{16}$$

$$U2 = \sqrt{\frac{\sum_{t=1}^{T}\left(\frac{f_t - RV_t}{HV_t}\right)^2}{\sum_{t=1}^{T}\left(\frac{HV_t - RV_t}{HV_t}\right)^2}} \tag{17}$$

For all metrics, smaller values indicate better fits, with a value of zero indicating a perfect fit. The usual Theil U2 statistic replaces HV_t with $RV_{t-1,22}$, and compares the given forecast with the most recent value

[2] For details please refer to the White Paper on the CBOE website (https://www.cboe.com/micro/vix/vixwhite.pdf).

of the forecasted variable. A value of unity indicates the forecast is no better than the naïve forecast of "no change." A value greater (less) than unity indicates the forecast is worse (better) than the naïve forecast. The reason for the modification in Equation (17) is that $RV_{t-1,22}$ is not observed at time $t-1$, and therefore cannot be used to forecast $RV_{t,22}$. The most recently observed value of realized volatility that can be used to forecast $RV_{t,22}$ is HV_t.

Further, the two predictors with the lowest RMSE are tested using the Diebold-Mariano ("DM") Test (Diebold-Mariano, 1995, Diebold, 2013). Broadly, the DM Test compares the errors of two different forecasts. This comparison is performed by defining an "error" as the difference between the projected value, and the actual value realized.

Determination of the Forecasting Regimes

Finally, in addition to investigating relative forecasting performances of the four forecasting variables, this study also investigates whether the relative performances differ across market conditions. The market conditions, or regimes, are distinguished by market returns and market volatility. The goal in choosing regimes was to have enough regimes to make for a meaningful analysis, but no so many that small quirks or idiosyncrasies within a data set could skew the results. A simple four regime model, which would contain a low-volatility, low-return, low-volatility, high-return, high-volatility low-return, and high-volatility, high-return was strongly considered, but the concern was that given the low level of regimes, that periods of extremely high volatility might skew the results.

The use of seven regimes, low-volatility, low-return, low-volatility, high-return, medium volatility, low-return, medium-volatility, high-return, high-volatility low-return, high-volatility, high-return, and an "all data" regime, which covered the entire sample period, was ultimately chosen for simplicity and to ensure that each regime had sufficient data to provide meaningful information. See Brownlees, Engle, and Kelly, (2011). Historical volatility was separated into the 33rd, 67th, and 100th percentile. Returns were separated into the 50th and 100th percentiles.

Accordingly, the seven regression regimes were created, as shown in Table 4.

Table 4

REGIME	VOLATILITY	RETURN
1	Low	Low
2	Low	High
3	Middle	Low
4	Middle	High
5	High	Low
6	High	High
7	All Data	

IV. RESULTS

The standard forecasting tests were then run for the four predictors for all seven regimes. The results indicate a high level of bias in many Regimes. A linear adjustment, explained on pages 14 and 15, was then made to help correct for the bias and the tests were re-run, with results indicating considerably less bias.

The results follow in Table 5:

Table 5

R1- Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.042	0.030	27.541	26.606	0.189	2.962
SGARCH	0.040	0.029	28.322	24.981	0.172	2.968
LGARCH	0.041	0.034	36.398	29.721	0.170	3.566
VIX	0.052	0.043	46.992	35.916	0.201	4.557

R2- Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.039	0.029	28.349	25.756	0.157	2.805
SGARCH	0.036	0.028	30.227	25.173	0.143	2.873
LGARCH	0.043	0.036	39.317	30.865	0.163	3.434
VIX	0.054	0.045	50.821	36.851	0.191	4.397

R3 - Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.069	0.048	30.410	29.672	0.201	2.484
SGARCH	0.065	0.044	28.621	26.520	0.188	2.437
LGARCH	0.059	0.043	30.397	26.755	0.168	2.478
VIX	0.072	0.059	46.578	35.502	0.185	3.579

R4 - Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.065	0.045	30.116	28.139	0.196	2.922
SGARCH	0.061	0.042	29.580	26.814	0.184	2.874
LGARCH	0.060	0.043	31.406	27.369	0.181	3.040
VIX	0.069	0.056	44.584	34.443	0.191	4.064

R5 - Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.086	0.059	31.164	27.393	0.181	3.118
SGARCH	0.080	0.053	28.360	25.323	0.171	2.720
LGARCH	0.071	0.045	24.863	22.697	0.152	2.382
VIX	0.093	0.072	42.619	33.387	0.184	3.575

R6 - Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.087	0.058	30.470	29.680	0.192	3.019
SGARCH	0.077	0.051	27.275	25.994	0.174	2.688
LGARCH	0.068	0.044	22.938	22.025	0.155	2.327
VIX	0.086	0.068	41.234	32.597	0.179	3.873

R7 - Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV	0.068	0.045	29.757	27.927	0.188	6.252
SGARCH	0.062	0.042	28.819	25.879	0.175	6.005
LGARCH	0.058	0.041	30.889	26.589	0.163	6.286
VIX	0.073	0.057	45.522	34.789	0.186	8.781

The Diebold Mariano Test was run using the two Predictors that had the highest R^2 (as shown in the regression in Table 7 on page 14.) The results follow in Table 6:

Table 6

Regime 1 - Diebold-Mariano test (HLN adjusted)				
VIX | LGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	15.07275	0.00	1.00	0.00
Sq Error	12.38521	0.00	1.00	0.00

Regime 2 - Diebold-Mariano test (HLN adjusted)				
VIX | SGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	22.03	0.00	1.00	0.00
Sq Error	18.25	0.00	1.00	0.00

Regime 3 - Diebold-Mariano test (HLN adjusted)				
VIX | LGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	14.269	0.000	1.000	0.000
Sq Error	8.732	0.000	1.000	0.000

Regime 4 - Diebold-Mariano test (HLN adjusted)				
VIX | SGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	12.962	0.000	1.000	0.000
Sq Error	5.486	0.000	1.000	0.000

Regime 5 - Diebold-Mariano test (HLN adjusted)				
SGARCH | LGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	7.722	0.000	1.000	0.000
Sq Error	5.633	0.000	1.000	0.000

Regime 6 - Diebold-Mariano test (HLN adjusted)				
VIX | LGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	20.744	0.000	1.000	0.000
Sq Error	14.727	0.000	1.000	0.000

Regime 7 - Diebold-Mariano test (HLN adjusted)				
VIX | LGARCH		Null hypothesis: Both forecasts have the same accuracy		
Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	39.454	0.000	1.000	0.000
Sq Error	25.499	0.000	1.000	0.000

In all Regimes, it appears that GARCH-based models outperform both historical volatility and VIX, as both, in all regimes, perform very poorly as forecasters of realized volatility. As shown in Figure 1, VIX very often overestimates realized volatility, and our data showed that VIX exceeded realized volatility over 85.0% of the time (Ge, 2016 found 84%). Thus VIX is an upwardly biased forecast of RV. This is further demonstrated by the regression of RV on VIX (See Table 7). These results might be puzzling at first glance, because the VIX is often thought of as the market's expectation of RV. Bekaert and Hoerova (2014) describe the VIX as follows: "The VIX index is the 'risk-neutral' expected stock market variance for the US S&P500 contract and is computed from a panel of options prices". However, the VIX is the risk-adjusted expectation of RV, not the actual expectation. Therefore these results are compatible with a negative volatility risk premium and are consistent with the findings of previous research (See Bakshi, Kapadia, 2003, Da, Schaumberg, November 21, 2011).

The extent of the bias can be measured by the coefficient on VIX; if this coefficient is equal to one, then VIX is an unbiased estimate. On the other hand, if this coefficient is much smaller than unity, VIX is more upwardly biased. Table 7 shows that in the entire sample, the coefficient is 0.893. In lower volatility regimes, the coefficient is as low as 0.52, showing a high level of upward bias. Interestingly, in the two high volatility regimes, the VIX coefficient as 1.10 and 0.90 respectively.

Specifically, In Regime 1, the low volatility, low return Regime, the standard GARCH model performed best, while historical volatility proved to be the best forecaster by two metrics the MAPE and the Theil U2. LGARCH significantly outperformed VIX in the Diebold-Mariano. In Regime 2, the SGARCH was again the best performer, but again the historical volatility outperformed in both the MAPE, and the Theil U2. GARCH models dominated the middle volatility regimes, and the LGARCH and SGARCH respectively outperformed the VIX via the Diebold-Mariano tests. LGARCH was also the strongest forecaster in the high volatility Regimes, outperforming all other Predictors in the standard tests, and outperforming the SGARCH and LGARCH respectively in the Diebold-Mariano. Consistent with expectations, the Theil U2 was exactly 1.00 for the Historical Volatility predictor.

The Diebold Mariano test further confirmed the weakness in using VIX as a Predictor of Realized volatility. In all instances comparing a VIX model to a GARCH based model, the results from both absolute error, and squared error showed positive values of a high magnitude, indicating the relative weakness in the VIX model. Specifically note that the probability that an alternative to VIX is a better predictor is at or close to one.

In an attempt to account for the bias in the forecasts, a series of linear adjustments were made to the Predictors. The first step was to run regressions on all predictors in all regimes, and use the results of these regressions as the basis for creating adjusted series.

The following are the results of the regressions run on the sample data.

Table 7

Ind. Variable	HV			SGARCH (1,1)			LGARCH(1,1)			VIX		
Regime	b_0	b_1	R^2	b_0	b_1	R^2	b_0	b_1	R^2	b_0	b_1	R^2
1	0.06	0.40	0.14	0.05	0.49	0.13	0.02	0.74	0.20	0.03	0.52	0.16
2	0.01	0.66	0.49	0.01	0.83	0.53	-0.01	0.88	0.46	0.00	0.76	0.57
3	0.06	0.67	0.39	0.04	0.79	0.40	0.00	0.95	0.49	-0.01	0.84	0.52
4	0.06	0.56	0.34	0.05	0.66	0.34	0.03	0.72	0.33	0.01	0.73	0.40
5	0.03	0.86	0.51	0.00	1.01	0.57	-0.01	1.06	0.66	-0.07	1.10	0.56
6	0.07	0.64	0.54	0.05	0.75	0.56	0.03	0.84	0.64	-0.02	0.90	0.57
7	0.04	0.73	0.53	0.02	0.86	0.55	0.00	0.96	0.60	-0.02	0.89	0.59

The regressions were then used to create the adjusted projection series for each Regime. The adjusted series were used to compute the "Adjusted Forecast" evaluation. For the adjusted forecasts, the results were adjusted linearly as follows:

$$RVA_t = b_0 + b_1 IndVar_t + \varepsilon_t$$

As an example, for VIX Regime 1, day 11, the following adjustment calculation was made:

$$RV_t = 0.0320 + (0.5234) * (0.225) = 14.96\%$$

The adjustments created four new series, as seen in Table 8:

Table 8

HV-A	Historical Volatility – Adjusted
SGARCH – A	Standard GARCH – Adjusted
LGARCH – A	Leverage GARCH – Adjusted
VIX – A	VIX – Adjusted

Then, the same standard forecasting tests were then run for the four predictors for all seven regimes. The results appear on the following page.

The following are the results using the adjusted variables:

Table 9

Regime 1	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.036	0.025	23.617	22.263	0.162	0.813
SGARCH-A	0.036	0.025	23.734	22.497	0.163	0.797
LGARCH-A	0.034	0.024	22.213	21.130	0.156	0.806
VIX-A	0.035	0.024	22.456	21.327	0.160	0.769

Regime 2	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.035	0.026	25.610	23.227	0.145	0.842
SGARCH-A	0.033	0.025	25.127	22.934	0.140	0.834
LGARCH-A	0.035	0.027	26.010	24.022	0.149	0.850
VIX-A	0.032	0.024	24.250	22.254	0.132	0.793

Regime 3	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.064	0.045	30.931	27.850	0.188	0.983
SGARCH-A	0.063	0.044	30.245	27.175	0.185	0.957
LGARCH-A	0.058	0.042	28.996	26.265	0.170	0.962
VIX-A	0.057	0.040	26.802	24.847	0.165	0.903

Regime 4	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.057	0.039	28.292	25.779	0.179	0.992
SGARCH-A	0.057	0.039	28.265	25.829	0.179	0.972
LGARCH-A	0.057	0.039	28.115	25.607	0.180	0.985
VIX-A	0.054	0.037	26.004	24.118	0.170	0.875

Regime 5	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.085	0.057	30.352	26.555	0.182	0.985
SGARCH-A	0.080	0.053	28.326	25.359	0.171	0.906
LGARCH-A	0.071	0.046	25.128	22.899	0.150	0.857
VIX-A	0.081	0.047	22.425	21.187	0.173	0.906

Regime 6	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.074	0.049	27.219	25.135	0.173	0.931
SGARCH-A	0.072	0.048	26.099	24.202	0.168	0.893
LGARCH-A	0.066	0.042	23.048	21.349	0.152	0.842
VIX-A	0.072	0.045	23.614	22.371	0.166	0.869

Regime 7	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
HV-A	0.063	0.042	28.956	26.162	0.181	0.913
SGARCH-A	0.061	0.040	28.113	25.540	0.175	0.885
LGARCH-A	0.058	0.039	27.674	25.208	0.165	0.890
VIX-A	0.059	0.037	24.738	23.093	0.168	0.849

For the adjusted Forecasts, the Diebold Mariano Test was run for the two predictors with the lowest RMSE results. The results are as follows:

Table 10

Regime 1 - A Diebold-Mariano test (HLN adjusted)

VIX - A | LGARCH - A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	0.804	0.422	0.789	0.211
Sq Error	1.709	0.088	0.956	0.044

Regime 2 - A Diebold-Mariano test (HLN adjusted)

VIX - A | GARCH -A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	-2.414	0.016	0.008	0.992
Sq Error	-3.434	0.001	0.000	1.000

Regime 3 - A Diebold-Mariano test (HLN adjusted)

VIX - A | LGARCH - A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	-2.671	0.008	0.004	0.996
Sq Error	-1.874	0.061	0.031	0.969

Regime 4 - A Diebold-Mariano test (HLN adjusted)

VIX - A | LGARCH - A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	-3.245	0.001	0.001	0.999
Sq Error	-4.357	0.000	0.000	1.000

Regime 5 - A Diebold-Mariano test (HLN adjusted)

VIX - A | LGARCH - A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	7.014	0.000	1.000	0.000
Sq Error	5.569	0.000	1.000	0.000

Regime 6 - A Diebold-Mariano test (HLN adjusted)

VIX - A | LGARCH - A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	3.893	0.000	1.000	0.000
Sq Error	4.456	0.000	1.000	0.000

Regime 7 - A Diebold-Mariano test (HLN adjusted)

VIX-A | LGARCH - A

Null hypothesis: Both forecasts have the same accuracy

Accuracy	Statistic	<> prob	> prog	< prob
Abs Error	-6.063	0.000	0.000	1.000
Sq Error	1.641	0.101	0.950	0.050

Discussion of Adjusted Regimes.

In the adjusted scenarios, the VIX performs much better relative to the other Predictors than in the base analysis, owing to the minimization of the impact of the variance risk premium. In Regime 1A LGARCH is minimally better via the standard tests, as well as per the Diebold-Mariano. In Regime 2A the VIX outperforms the LGARCH in both standard tests and in the Diebold-Mariano. In Regime 3A VIX has a di-minimis edge in the standard tests, and outperformed LGARCH by a low magnitude in the Diebold-Mariano. In Regime 4A VIX outperformed LGARCH by a slightly greater level in both standard tests and the Diebold-Mariano. In the higher volatility regimes, Regime 5 and Regime 6, the leverage GARCH was a stronger predictor. In Regime 5, LGARCH strongly outperformed SGARCH in all tests, including the Diebold-Mariano. Note too, that although the SGARCH was chosen over the VIX based upon the lower RMSE, the VIX actually outperformed the LGARCH in the MAPE, the SMAPE, and the Theil U2, and performed relatively better than the SGARCH when compared to the LGARCH in the Diebold-Mariano. In Regime 6, LGARCH was a stronger predictor in all but the Thiel U2. Interestingly, in Regime 7, The LGARCH outperformed the VIX in squared statistics, including the squared error in the Diebold-Mariano, but was outperformed by the VIX in the other statistics.

V. CONCLUSION

There seems to be some indication that the Volatility/Return regime, as defined herein, does impact the efficacy of methods used to predict volatility. Notably none of the methods appeared to be particularly effective in predicting volatility in low-volatility, low-return periods. It was noted that the leveraged GARCH demonstrated significantly less bias than VIX models in low volatility periods, perhaps owing to a much higher volatility risk premium in low volatility environments, however the VIX model appears marginally less biased in the high volatility regimes.

It is less clear that the Predictors have different relative capabilities in different regimes, however the results from Regimes 5 and 6 appear to indicate that in a high-volatility environments the leveraged GARCH model may be a superior predictor.

The volatility risk premium also seems to be impacted by the regime. In the two low volatility regimes, the coefficient was at 0.52 and 0.76, while in the high volatility regimes, was at 1.1 (indicating a negative volatility risk premium), and 0.90

Further work will involve the use of an out of sample study, an attempt to determine other factors that might contribute to creating better predictors while in low volatility scenarios, and an attempt to better identify the impact of the volatility risk premium in preparing this predictive analysis. It also would be useful to expand the number of regimes, likely expanding the number of volatility cutoffs, rather than return cutoffs, given the apparent lesser impact of the return component.

Further development of a more definitive understanding of how various regimes impact the efficacy of various predictors could meaningfully alter the approach taken by portfolio managers, risk managers, and volatility traders in predicting volatility. Managers may start to use different models in different regimes, or possibly more heavily weight one approach over the other in constructing their volatility models based upon the regime. One of the key burdens in any attempt to use this finding will be the difficulty in identifying precisely what volatility return regime we are a priori at any particular time.

REFERENCES

Bakshi, Gurdip, Kapadia, Nikunji, Delta-Hedged Gains and the Negative Market Volatility Risk Premium, *The Review of Financial Studies*, Volume 16, Issue 2, 1 April 2003, Pages 527–566,

Bekaert, G., Hoerova, M., (May 2014). The VIX, the Variance Premium and Stock Market Volatility. NBER Working Paper Series, National Bureau of Economic Research

Bollerslev, T. Generalized Autoregressive Conditional Heteroskedasticity GARCH paper

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H., Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence. Journal of Financial and Quantitative Analysis Vol. 49, No. 3, June 2014, pp. 633–661.

Brownlees, C., Gallo, G. Comparison of Volatility Measures: a Risk Management Perspective Journal of Financial Econometrics, 2010. Vol 8, No. 1, 29-56

Campbell, J., Lo, A., MacKinley, A., (1997), The Econometrics of Financial Markets, Princeton University Press.

CBOE, The CBOE Volatility Index – VIX® https://www.cboe.com/micro/vix/vixwhite.pdf

Da, Zhi, Schaumberg, Ernst, The Pricing of Volatility Across Asset Classes, November 21, 2011, Mendoza College of Business, University of Notre Dame.

Department of Treasury, Forecasting Accuracy of the ACT Budget Estimates, May 2008

Diebold, F.X. and R.S. Mariano. (1995). Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13: 253-63.

Diebold, F., (December 5, 2013), Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests, Journal of Business and Economic Statistics Invited Lecture, Allied Social Science Association Meetings, Philadelphia, PA.

Fama, E., French, K., (June, 1992), The Cross-Section of Expected Stock Returns, The Journal of Finance.

Ge, W., (January, 2016),Understanding the Sources of the Volatility Risk Premium, Parametric Portfolio Associates, LLC.

Littner, J., (February, 1965), The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgeting. CAPM paper. The Review of Economics and Statistics, Vol. 47, No. 1, pp. 13-37

Poon, S, Granger, C., (June 2003), Forecasting Volatility in Financial Markets: A Review, Journal of Economic Literature, Vol. XLI pp 478-539.

Sharpe, W., (September, 1964), Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk, The Journal of Finance, Vol. XIX, No. 3.

Todorov, V. Variance Risk-Premium Dynamics: The Role of Jumps, Kellogg School of Management, The Review of Financial Studies / v 23 n 1 2010