International Review of Business Research Papers
Vol. 14. No. 1. March 2018 Issue. Pp. 205 – 225

Forecasting Power of Five CBOE Volatility Indexes for the Price Interval

Young Cheol Jung*

This paper evaluates the forecasting power of five CBOE Volatility Indexes (VIX; VXN; OVX; EVZ; GVZ) for the price interval. For the period of June 2008- March 2017, this paper builds a series of one-month ahead price intervals of underlying asset with a band of ± 1 or 2 times of its daily volatility index. Then assesses whether each volatility index is a good estimator for the price interval through comparing its performance to that of the ex-post realized volatility and normal distribution benchmark. Further this paper tests whether the expected shortfall (ES) beyond the price interval forecast of the volatility index is lower than ES of the realized volatility and the benchmark generated by Monte Carlo Simulation. Based upon the test results, this paper suggests that VIX and VXN are robust estimators for the future price interval and can be used in estimating the Value-at-Risk (VaR). VIX and VXN performed well particularly in predicting the upper bound of returns due to the feature of the asymmetric negative return/volatility index relation that exists in the stock markets. The evaluation for OVX, EVZ and GVZ is reserved because their performance was inconsistent and inferior to VIX and VXN.

JEL classification: C22, C53, G17

1. Introduction

In the financial industry, the price volatility plays an important role as a key input for pricing the derivatives, allocating assets and managing risk. Due to its importance, testing the forecasting power of volatility estimators has held attention and challenge over the last three decades. However, until recently, the focus of the test, has been paid to a point forecast- computing a single “best guess” of volatility and comparing it to the realized volatility and/or other volatility estimators. More precisely, the previous studies on the predictability of volatility index (or implied volatility) tested its forecasting power for realized volatility- a proxy for “true” volatility - or model-based volatilities (e.g., the GARCH volatilities), mainly by using two standard methods complementarily: the OLS regressions and/or the loss functions. The previous studies also commonly tried to rank volatility estimators according to how far a volatility estimator is deviated from the realized volatility. However, due to conflicting results, any “consensus” on the ranking could not be fully reached (Figlewski 1997; Poon & Granger 2003 for extensive survey on volatility forecasting).

Even though the point forecast still has first-order importance, it cannot be denied that it is subject to several limitations.[1] First, the “true” volatility- a test benchmark- is not observable. In the previous literature, three methods have been commonly used to find a proxy for the “true” volatility of financial asset; i.e., the ex-post sum of the daily squared returns (in most early

* Young Cheol Jung, CFA, Assistant Professor, Department of Economics, Justice, and Policy Studies, Mount Royal University, 4825 Mount Royal Gate SW, Calgary, Alberta, Canada, T3E 6K6, 403-440-5079, Email: yjung@mtroyal.ca

literature) or the sum of the intraday squared returns (Taylor & Xu 1997; Andersen & Bollerslev 1998; Martens & Zein 2004) or the daily range-based estimator (Garman & Klass 1980; Yang & Zhang 2000; Kristoufek 2014). Although the sum of the intraday squared returns using high-frequency data is recently most popularly used, it is still far from any clear picture of the “true” volatility because the overnight and over-the-weekend’s return cannot be measured with the same frequency (Martens 2002; Koopman et al. 2005). Second, different users will have different interests in volatility. While a more accurate point forecast for volatility is profitable for option or variance swap investors, the point forecast is not so informative for portfolio (e.g., ETF) investors- whose main concern is not the volatility itself but a downside or upside risk. In this context, the prediction for tail distributions is of particular importance for the portfolio investors. To address the issue, this paper focuses on assessing the forecasting power of volatility indexes in regards to the tail distribution of returns rather than the point forecast. This different focus is expected to help dodging the measurement problem of the “true” volatility and further, providing the portfolio investors with an additional useful information.

Since the Chicago Board Options Exchange (CBOE) first introduced VIX in 1993, which measures 30-day expected volatility of the S&P500 index, the volatility indexes became a popular vehicle for forecasting market volatility. VIX estimates the realized volatility by averaging the weighted prices of SPX put and call options over a wide range of strike prices. Using the VIX algorithm, CBOE has developed more than 25 volatility indexes. Considering the data availability and its representativeness, this paper chose five volatility indexes- VIX, VXN, OVX, EVZ and GVZ, which are overviewed in Table A1 (Appendix).[2]

It is generally recognized that the volatility index can be used to estimate the expected price range of its underlying asset at a specific point of time – 30 days ahead. For example, in June 2, 2008, VIX and SPX was 19.83 and 1385.67, respectively. Then SPX is expected to be between 1308.58 and 1467.31 at 68.3% level and between 1235.77 and 1553.75 at 95.4% level in July 1, 2008 (21 trading days or equivalently 30 calendar days later). In this regard, the first question of this paper is whether the 21 trading days- ahead prices of underlying asset has been within the price range implied by the volatility index at each confidence level. The second question is whether the accumulated density of the deviations beyond the expected price range is smaller than that of benchmark.

Whereas numerous research tested the point forecasting power of volatility indexes, there have been few studies on their predictability for the price interval at a given confidence level. Furthermore, the studies on the forecasting power of volatility indexes for non-stock market are rare.[3] This research is different from previous studies in the sense that it is a comparative study on the forecasting power of five CBOE volatility indexes for the price range. In this perspective, the findings of this research is expected to fill the void of the previous studies.

The rest of this paper proceeds as follows. In the next section, the relevant literature is reviewed. Section3 presents the data source, methodology for the data transformation and descriptive statistics. Section4 describes the test methodology for the price interval and ES and reports the test results. The last section summarizes and concludes.

2. Literature Review

Traditionally, tests for forecasting power have concentrated on methods to evaluate the point forecast. Although the point forecast is important, it has a limited value due to its only one side description of possible outcome.[4] Recently more attention is paid to the tail distribution of returns, which is composed of two dimensions- the price interval and density. The value-at-risk (VaR) - i.e., a minimum loss of a financial position within a certain period of time at a given probability - is the most popular measure for the price interval (see Morgan 1996).

There have been several influential papers on the forecast for the price interval. For example, Chatfield (1993) was the first to review the importance of the interval forecast by describing and comparing several approaches to calculate the interval forecast. Christoffersen (1998) introduced a general conditional efficiency criterion for evaluating the interval forecast and emphasized that there is more information in interval forecast than point forecast. Diebold et al. (1998) developed a framework to evaluate the density forecasts for financial risk management. Giot (2005) was the first study on the applications of volatility indexes of the US stock market – VXO, VXN and VIX- to estimate VaR and showed that the volatility index scaled for one-day provided meaningful volatility information in VaR models.

The accuracy of the price interval predicted by VaR can be assessed by back-testing whether the actual number of “exceedance”- i.e., an incident of a portfolio’s loss exceeding its VaR - is smaller than the expected number of violations in the normal distribution at the same confidence level. As for one of the back-testing methodologies to evaluate the statistical accuracy of the price interval forecast, Kupiec (1995) proposed the unconditional coverage test to analyze the statistical significance of the observed frequency of violations with respect to the expected one. Christoffersen (1998) suggested the independence test to gauge the independence of violations, i.e., whether the likelihood of a VaR violation today depends on the VaR violation of the previous day; the conditional coverage test combines these two tests.[5]

Although VaR is a popular measure of tail risk, it does not offer any insight into the potential size of the loss in excess of the VaR level. Consequently, the interval forecast cannot capture the probability distribution of the portfolio returns beyond the predicted price interval. Moreover, as shown firstly by Artzner et al. (1997, 1999), VaR is not a coherent risk measure because it doesn’t fulfil one of the axioms of coherence.[6] As a natural coherent alternative to VaR, Acerbi and Tasche (2002a & b) pointed out that ES - i.e., the average of losses beyond the price interval estimated at a given confidence level- yields a coherent risk measure regardless of the underlying distributions and prosed it as remedy for the deficiencies of VaR. Hence ES provides VaR with additional information about the tail distribution of returns. In that sense, ES can be considered as one of the density forecast measurements for extreme value.

The density forecast is defined as an estimate of the entire probability distribution of the possible future values of a variable (Tay & Wallis 2000). Despite a practical importance of the density forecast for portfolio investors, there has been relatively scarce research on testing the density forecast. Among the previous studies, Diebold et al. (1998) considered the conditional dynamics of the density forecast by examining the correlograms and powers of the probability integral transformations and reported an application to density forecast of S & P 500 returns. McNeil and Frey (2000) introduced the methods based upon extreme value theory to improve the estimation of the tail distribution and estimated the conditional ES by using the generalized Pareto distribution (GPD) function. Berkowitz (2001) combined the forecast distribution with ex-post realization to construct testable hypothesis even with sample sizes as small as 100. One disadvantage of ES is that ES estimates are less accurate than VaR estimates when the losses have fat tails. Thus it requires to reduce the estimation error of ES by increasing sample size (Yamai & Yoshiba 2002, 2005). Another disadvantage of ES is that ES is not “elicitable” (Gneiting 2011) and thus back-testing ES is less straight-forward than VaR (Emmer et al. 2015). Regarding the back-testing methods, Kerkhof and Melenberg (2004) back-tested VaR and ES using the functional delta method and indicated that, contrary to common belief, ES is not harder to back-test than VaR if the level of ES is adjusted. Wong (2008) used the saddlepoint technique to back-test the trading risk of commercial banks using ES and showed that the back-test is accurate and powerful even for small samples. Costanzino and Curran (2015) presented the quantile-space coverage test for spectral risk measure. In the most comprehensive way, Acerbi and Szekely (2014) introduced three model-free, nonparametric back-test methods for ES, observed that elicitability has nothing to do with model testing and concluded that ES can be back-tested. More empirically, Pedersen (2015) used quantile regression to examine predictability for the left tail, center and right tail of the distribution and compared the relative performance to the normal distribution by using the weighted likelihood ratio test (see also Amisano & Giacomini 2007).To assess the predictability of implied volatility for VaR, Cesarone and Colucci (2016) proposed a model to forecast ex-ante VaR using combined estimator between realized volatility and implied volatility and confirmed that the efficacy of the implied volatility as inputs for VaR. Kim and Ryu (2015), on the other hand, indicated that VKOSP- the volatility index for KOSPI200- did not enhance the performance of VaR models compared to other volatility forecasting models.

However, the previous literature on forecasting the interval by using VaR model is mainly confined to the one-day VaR (e.g., Giot 2005; Rubia & Sanchis-Marco 2013) or the downside risk (e.g., McNeil & Frey 2000; Berkowitz 2001; Yamai & Yoshiba 2005; Rubia & Sanchis-Marco 2013). In particular, the back-testing methods for ES (e.g., Berkowitz 2001; Kerkhof & Melenberg 2004; Wong 2008) are difficult to be implemented for the financial markets of which return shows non-Gaussian distribution (Chen & Gerlach2013). Moreover, there has been few research for testing the accuracy of the price range predicted by the CBOE’s volatility indexes (e.g., Giot 2005; Haugom et al. 2014; Luo et al. 2016).

Given the aforementioned limitations of previous literature, this paper adopts an existing methodology but revises it in a more practical and implementable way. Also this paper uses the market indexes of the five CBOE underlying asset to test the predictability of its volatility index for the one-month price range rather than one-day price range. Further, the predictability is tested for both upside and downside of the price range. For the purpose, first, this paper estimates VaR of each underlying asset, which is implied by the volatility index and then, as suggested by Kupiec (1995), the forecasting power of volatility indexes for the price interval is assessed through a binominal test on the observed frequency of violations from VaR. To backtest ES beyond VaR, each series of ES is resampled by applying a Bootstrapping method to approximate the finite-sample distribution of ES (see Chatfield 1993; Mancini & Trojani 2011) and the Bootrapped ES are compared to standardized density of the normal distribution.

3. Data and Methodology

The data consist of the daily closing price (2224 observations) of five volatility indexes and their corresponding underlying assets spanning the period of June 2008 - March 2017 (106 months). The sample includes periods of high turbulence such as the global financial crisis, the euro sovereign debt crisis, the rebound in the US stock market and the market turmoil of gold and the crude oil.[7] To assess the price interval forecasting power of volatility indexes (IV), first of all, it needs to identify the "true" volatility (i.e., standard deviation) as the natural benchmark. This paper adopts the realized volatility (say, RV) as a proxy for the unobservable "true" volatility. Despite several methods to estimate RV, this paper calculates it from the overlapping data by assuming zero sample mean and then annualizing the average volatility of 21 trading days by multiplying the square root of 365 as below.[8]

$$RV_{t,t+21} = \sqrt{\frac{365}{21}\sum_{j=1}^{21}\left(\ln(p_{t+j} / p_{t+j-1})\right)^2} \times 100 \qquad (1)$$

where p_t : a closing index at day t.

Table1 summarizes the descriptive features of IV and RV. In most previous studies, it is observed that the average of volatility index (IV) is greater than that of the realized volatility (RV) due to the volatility risk premium (e.g., Corrado & Miller 2005; Carr &Wu 2006). However, Table1 exhibits that IV was even a little lower than RV except for SPX. The conflicting result of this paper is due to the multiplication of 365 day to annualize RV in order to maintain the consistency with the CBOE's calculation method for IV. Among the five underlying asset, USO was the most volatile (i.e., RV=39.09, IV=38) while FXE was the least volatile (RV=11.93, IV=11.54). Due to right skewness and high kurtosis, the null hypotheses of normal distribution are rejected for all. Knowing that the descriptive features were similar between IV and RV of each underlying asset, it will be intriguing to compare the forecasting power of each IV and RV.

Table 1: Descriptive Statistics for IV and RV (June 2008 - March 2017)

		Underlying Index				
		SPX	NDX	USO	FXE	GLD
IV	Mean (a)	20.70	22.05	38.00	11.54	21.14
	Maximum	80.86	80.64	100.42	30.66	64.53
	Minimum	10.32	10.31	14.5	4.69	10.8
	Std. Dev.	10.18	9.82	14.76	3.88	7.64
	Skewness	2.31	2.33	1.19	1.32	2.20
	Kurtosis	9.53	9.92	4.98	5.67	8.95
	Jacque- Bera	5930.12	6447.76	883.83	1302.24	5074.95
	P-value	.000	.000	.000	.000	.000
RV	Mean (b)	20.54	22.52	39.09	11.93	21.32
	Maximum	101.34	101.91	109.60	30.22	74.00
	Minimum	4.25	1.23	11.09	2.03	4.03
	Std. Dev.	15.00	14.20	19.18	4.58	9.72
	Skewness	2.78	2.75	1.28	1.12	2.02
	Kurtosis	12.35	12.68	4.65	4.92	8.41
	Jacque- Bera	10965.63	11484.64	860.6	804.66	4221.64
	P-value	.000	.000	.000	.000	.000
a-b		+.16	-.47	-1.09	-.39	-.18

The experiment of this paper is to test the null hypothesis that IV cannot be used in predicting the price interval or VaR of its underlying asset i. Given some confidence level $\mathrm{q} \in (0,1)$, VaR of a portfolio at the level q is defined to be the smallest number x such that the probability of loss L to exceed x is no greater than $(1-\mathrm{q})$. VaR is expressed:

$$x_q = inf\{x_q \in R : p(L \geq x_q) \leq 1 - q\} \qquad (2)$$

The method to test the hypothesis is practical rather than theoretical so that it is easily implementable for back-testing the forecasting power of the volatility indexes for the price interval. If the CBOE volatility index can be used for the analytical VaR method, IV will represent the upper or lower bound of the expected % change in price of its underlying asset over the month at 68.3% confidence level. T-day ahead price interval ($p_{i,t+T}$) of a financial asset *i* at (1-α)% confidence level is given at time *t* by the following general form:

$$p_{i,t} \pm z_{\alpha/2}\,\hat{\sigma_i} \qquad (3)$$

where$z_{\alpha/2}$ denotes the z-score of a normal distribution and $\hat{\sigma_i}$ is the estimated volatility of the asset *i*.

Table 2: Descriptive Statistics of NDEV for IV and RV: $\tau = 1$ (68.3%) (June 2008 - March 2017)

		Underlying Index				
		SPX	NDX	USO	FXE	GLD
IV	Mean (a)	.5444	.5805	.3948	.4259	.5179
	Maximum	1.3469	1.4529	1.5518	1.8752	2.3472
	Minimum	-1.5049	-1.3195	-.9844	-1.2046	-1.5112
	Std. Dev.	.3707	.4007	.441	.4508	.4494
	Skewness	-1.18	-.95	-.3	-.30	-.04
	Kurtosis	5.29	4.26	2.72	3.36	3.74
	Jacque- Bera	1001.24	481.75	41.57	45.04	51.12
	P-value	.000	.000	.000	.000	.000
RV	Mean (b)	.6290	.6646	.4346	.4432	.5329
	Maximum	1.7951	1.7737	1.6511	1.6473	1.8621
	Minimum	-.5382	-.4024	-.6752	-.8048	-.7392
	Std. Dev.	.3787	.4084	.4204	.4156	.4270
	Skewness	.05	.13	.13	-.03	.05
	Kurtosis	2.59	2.52	2.25	2.63	2.68
	Jacque- Bera	16.82	27.31	58.37	13.10	10.74
	P-value	.000	.000	.000	.001	.005

Table 3: Descriptive Statistics of NDEV for IV and RV: $\tau = 2$ (95.4%) (June 2008 - March 2017)

		Underlying Index				
		SPX	NDX	USO	FXE	GLD
IV	Mean (a)	.5000	.5166	.4068	.4505	.4861
	Maximum	.8909	.9555	.9991	1.1742	1.4041
	Minimum	-.5263	-.4364	-.2673	-.3672	-.5186
	Std. Dev.	.1857	.2008	.2196	.2249	.2244
	Skewness	-1.18	-.94	-0.29	-.30	-.01
	Kurtosis	5.25	4.24	2.69	3.36	3.73
	Jacque- Bera	980.14	468.55	40.66	44.24	50.00
	P-value	.000	.000	.000	.000	.000
RV	Mean (b)	.5425	.5581	.4258	.4587	.4934
	Maximum	1.1401	1.1240	1.0224	1.0624	1.1607
	Minimum	-.0480	.0176	-.1195	-.1624	-.1319
	Std. Dev.	.1950	.2100	.2136	.2078	.2142
	Skewness	.03	.11	.14	-.03	.07
	Kurtosis	2.56	2.50	2.25	2.62	2.66
	Jacque- Bera	18.20	27.42	59.72	14.05	12.52
	P-value	.000	.000	.000	.001	.002

This paper calculates the price interval of an underlying asset with a more familiar convention; i.e., rather than calculating VaR, this paper uses a band of $\pm$ one or two standard deviation ($\pm\sigma, \pm 2\sigma$) around the point forecast of estimated volatility. Then the T-day ahead lower bound $\left(p^{e}_{l,i,t+T}\right)$and upper bound $\left(p^{e}_{h,i,t+T}\right)$ of an underlying asset *I* are predicted at time *t*:

$$p^{e}_{l,i,t+T} = p_{i,t}e^{\frac{-\tau V_{i,t}/\sqrt{12}}{100}}$$
$$p^{e}_{h,i,t+T} = p_{i,t}e^{\frac{+\tau V_{i,t}/\sqrt{12}}{100}} \quad (4)$$

where $V_{i,t}$: estimated volatility, T: ahead-of-time (e.g., T=21 for 1-month), τ: scaling factor for confidence level (i.e., τ = 1 for 68.3% and τ = 2 for 95.4%)

Since the level of each volatility index is different, this paper calculated the normalized deviation (NDEV) – i.e., the ratio of the actual 21 trading day (i.e., approximately 30 calendar days) ahead deviation from the expected lower bound to the expected maximum price range as shown in the equation (5):

$$NDEV_{i,t+21} = \frac{p_{i,t+21} - p^{e}_{l,i,t+21}}{p^{e}_{h,i,t+21} - p^{e}_{l,i,t+21}} \quad (5)$$

If $0 \leq NDEV \leq 1$ holds, the underlying asset price at t+21 is within the predicted price interval at the predetermined confidence level. In that case, the mean (μ) of *NDEV* is expected to be 0.5 at both confidence levels and the standard deviation (σ) of *NDEV* is expected to be 0.5 at τ = 1 (68.3% confidence level) and 0.25 at τ = 2 (95.4% confidence level). *NDEV*= 0 and 1 correspond to the cut-off for the left and right tail of VaR, respectively, at both confidence level.

Table 2 and 3 demonstrate the statistical features of NDEV for each IV and RV. At first glance, it is found that the means of NDEV are the highest for NDX and the lowest for USO, which can be interpreted that the market was the most bullish in NDX and the most bearish in USO during the sample period. Also it is observed that the data transformation into NDEV considerably reduced skewness, kurtosis and Jackque-Bera of each estimator compared to the results in Table 1 even though the distributions of NDEV are still non-normal. Several research found that the VaR estimate obtained under skewed and fat-tailed distributions provides a more accurate VaR than those obtained from a normal distribution (e.g., Abad et al., 2014). With this justifications, this paper will use the transformed data (i.e., NDEV) for the comparison with a benchmark of normal distribution.

4. Results

4.1. Price Interval

To test the forecasting power of IV and RV for the price interval, the actual number of violation (v_i) of a portfolio*i* is calculated at each confidence level ($\pm\sigma, \pm 2\sigma$) by using 21-day ahead NDEV with *n* observations:

$$(\text{Right tail})\, \upsilon_i = \sum_{t=1}^{n} I_{i,t+21}, I_{i,t+21} = 1 \; if \; NDEV_{i,t+21} > 0 \quad (6)$$
$$= 0 \; if \; otherwise$$
$$(\text{Left tail})\, \upsilon_i = \sum_{t=1}^{n} I_{i,t+21}, \;\; I_{i,t+21} = 1 \;\; if \; NDEV_{i,t+21} < 0$$
$$= 0 \; if \; otherwise$$

Next the actual number of violations (v_i) is compared with the expected number of violations ($\hat{v}$) of a normal distribution. Kupiec (1995) showed that the probability of violation is constant, the number of violation follows a binomial distribution. Therefore, this paper proceeds with one-side binomial test for the number of violations to check if IV or RV performed better than its benchmark.[9] For the binomial test, the following hypotheses were set up:

$$Ho : v_i \geq \hat{v} \qquad (7)$$
$$H_A : v_i < \hat{v}$$

If the p-value of the test statistic is less than 0.1, 0.05, 0.01, an alternative hypothesis is accepted at 10%, 5%, and 1% significance level, respectively. The acceptance of the alternative hypothesis signifies that there was a less violation than expected, which satisfies a condition to be a good estimator for price interval. The violation ratio (V-Ratio), which is defined the actual number of violations to the expected number of violations beyond the predicted price interval, is used to compare the forecasting power for the price interval between volatility estimators.

Table 4 and 5 report the test results on the price interval forecast of IV and RV. In Table 4, the binominal tests confirm that all IVs are good estimator for the price interval in the right tail: i.e., the actual number of violations were all smaller than the expected number of violations at both intervals ($+\sigma, +2\sigma$). IVs of SPX, NDX and USO were even superior to RV in the right tail. In particular, the volatility indexes for stock market- VIX and VXN were outstanding and accomplished less number of violations than expected in both tails and both confidence levels. However, as shown in Table 5, IVs of USO and FXE performed disappointingly in forecasting the price interval of the left tail. The poor performance of OVX and EVZ for the left tail prediction may be related to recent plunge of the crude oil price and long-lasting slump of euro currency since the global financial crisis in 2008-9. IV of GLD (i.e., GVZ) achieved less number of violation than expected for both tails but its forecasting power for the price interval was not better than RV's in most cases.

Forecasting the price interval, however, has a limited assessment on the risk because the tail distribution of loss and gain cannot be properly evaluated. As aforementioned, VaR only offers a bound on the losses that occur with a confidence level so that it tells nothing about the potential size of the losses exceeding the bound. More formally, VaR at confidence level q does not give any information about the severity of losses which occur with a probability less than 1-q. As a remedy for the limitations of VaR, the expected shortfall (ES) is a density measure to be used complementarily with VaR by providing portfolio investors with the magnitude of downside or upside risk.ES is formularized as the expected loss (L) conditional on the loss being above the VaR level (x_q):

$$y_q = E\{L \middle| L \geq x_q\} \qquad (8)$$

ES is sensitive to the size of losses beyond VaR because it considers the full distribution of tail. Once the loss exceeds the VaR threshold, ES better quantifies the loss. Furthermore, since ES is a coherent risk measure (Acerbi & Tasche 2002b), it allows a more straightforward allocation of capital and hence reasonable to use for determining capital requirement (Costanzino & Curran 2015). In this context, the second experiment of this paper is to conduct a supplementary test to investigate whether ES as a consequence of the price interval forecast by the volatility index is lower than ES of RV and the benchmark.

Table 4: Backtesting Results for Price Interval (Right Tail) (June 2008 - March 2017)

Interval	# of violation		Underlying Index				
			SPX	NDX	USO	FXE	GLD
+σ	Expected (e)		353				
	Actual	IV (a)	120***	281***	152***	204***	282***
		RV (b)	368	466	240***	217***	340
	V-Ratio	a/e	**.340**	**.796**	**.431**	**.578**	**.800**
		b/e	1.043	1.321	.680	.615	.964
+2σ	Expected (e)		51				
	Actual	IV (a)	0***	0***	0***	12***	37**
		RV (b)	12***	37**	2***	6***	18***
	V-Ratio	a/e	**.000**	**.000**	**.000**	.237	.732
		b/e	.237	.731	.040	**.119**	**.356**

Table 5: Backtesting Results for Price Interval (Left Tail) (June 2008 - March 2017)

Interval	# of violation		Underlying Index				
			SPX	NDX	USO	FXE	GLD
-σ	Expected (e)		353				
	Actual	IV (a)	175***	202***	441	366	276***
		RV (b)	104***	99***	386	337	224***
	V-Ratio	a/e	.496	.572	1.250	1.037	.783
		b/e	**.295**	**.281**	**1.094**	**.955**	**.635**
-2σ	Expected (e)		51				
	Actual	IV (a)	38**	41*	80	83	25***
		RV (b)	2***	0***	9***	21***	16***
	V-Ratio	a/e	.751	.810	1.581	1.640	.494
		b/e	**.040**	**.000**	**.178**	**.415**	**.316**

Note: 1. A better interval estimator is boldfaced in the column (4)-(8).
2. *, **, *** denotes significance at the 10%, 5% and 1% level, respectively to accept the alternative hypothesis.

4.2. Expected Shortfall (ES)

To further assess the forecasting power for the tail distribution, this paper estimated the expected shortfall ($\widehat{\mathrm{ES}}_b$) of a benchmark by carrying out the Monte Carlo experiment. Through the simulation, a sequence of 2,224 (i.e., number of sample) NDEV is repeated 10,000 times with the distribution of $NDEV_b \sim \mathrm{N}(\mu, \sigma^2) = \mathrm{N}(0.5, 0.5^2)$ for $\pm\delta$ and $\mathrm{N}(0.5, 0.25^2)$ for $\pm 2\delta$. $\widehat{\mathrm{ES}}_b$ is calculated by equation (9):

(Right tail) $$\widehat{\mathrm{ES}}_b = \frac{\sum_{t=1}^{n} NDEV_{b,t+21} I_{b,t+21}}{\sum_{t=1}^{n} I_{b,t+21}} \; where I_{b,t+21} = 1 \;\; if \; NDEV_{b,t+21} > 1 \quad = 0 \; if \; otherwise \qquad (9)$$

(Left tail) $\widehat{ES}_b = \frac{\sum_{t=1}^{n} NDEV_{b,t+21} I_{b,t+21}}{\sum_{t=1}^{n} I_{b,t+21}} where I_{b,t+21} = 1 \ if \ NDEV_{b,t+21} < 0$
$= 0 \ if \ otherwise$

The numbers in the 3rd and 8th row of table 6 and 7 are the benchmark's expected shortfall ($\widehat{ES}_b$) beyond the price interval of $\pm\sigma$ and $\pm 2\sigma$. The actual shortfall (ES_i) of a portfolio *i* is calculated at each intervalby applying the equation (9). ES estimates are less accurate than VaR estimates when the losses have fat tails (Yamai & Yoshiba 2002). Since the $NDEV$ s are not normally distributed (see Table 2 and 3), this paper used Bootstrapping methods to estimate the ES in order to reduce the estimation error by increasing sample size.

To evaluate the loss beyond the price interval, the bootstrapped ES_i in each of short (right tail) and long (left tail) position was compared with the benchmark, $\widehat{ES}_b$. One-side *t*- test is formulated as below to check whether IV or RV outperformed (i.e., smaller ES in an absolute value) its benchmark in terms of ES:

(Right tail) $Ho: ES_i \geq \widehat{ES}_b$ vs. H_A: $ES_i < \widehat{ES}_b$ (10)

(Left tail) $Ho: ES_i \leq \widehat{ES}_b$ vs. H_A: $ES_i > \widehat{ES}_b$

If the p-value is less than 0.1, 0.05, 0.01, an alternative hypothesis is accepted at 10%, 5%, 1% significance level, respectively, which implies that ES_i is less than $\widehat{ES}_b$ in an absolute value. The acceptance of the alternative hypothesis satisfies a condition to outperform the benchmark. In Table 6 and 7, the 4th and 9th row of column (4)-(8) report ES_i when each price interval was estimated by IV. The 5th and 10th row present ES_i if RV was used to estimate the price interval (although it is an impossible assumption). In Table 6, IVs for SPX, NDX and USO still performed best in the right tail: i.e., their ES is lower than ES of RV and benchmark at both confidence levels. IV of FXE also performed well for the right tail at 68.3% confidence level. However, as shown in Table 7, none of IVs statistically have achieved better than RV and benchmark in the left tail. Even though OVX outperformed the benchmark in the right tail, its performance is disappointing in the left tail. IV of GLD achieved a good performance for the price interval, but it did not show any discernible outcome for ES. IV of FXE (i.e., EVZ) also displayed a weakness in the left tail forecast.

Table 6: Backtesting Results for ES (Right Tail)
(June 2008 - March 2017)

Interval	ES		Underlying Index				
			SPX	NDX	USO	FXE	GLD
+σ	Expected (e)		1.2626				
	Actual	IV (a)	1.0925***	1.1111***	1.1384***	1.1848***	1.2549
		RV (b)	1.2018***	1.2458	1.1501***	1.1591***	1.1986***
	V-Ratio	a/e	**.8652**	**.8800**	**.9016**	.9383	.9939
		b/e	.9519	.9867	.9108	**.9180**	**.9493**
+2σ	Expected (e)		1.0932				
	Actual	IV (a)	.0000***	.0000***	.0000***	1.0846	1.0933
		RV (b)	1.0352	1.0378*	1.0212	1.0374	1.0528
	V-Ratio	a/e	**.0000**	**.0000**	**.0000**	.9921	1.0001
		b/e	.9469	.9493	.9341	**.9489**	**.9630**

Table 7: Backtesting Results for ES (Left Tail)
(June 2008 - March 2017)

Interval	ES		Underlying Index				
			SPX	NDX	USO	FXE	GLD
-σ	Expected (e)		-.2625				
	Actual	IV (a)	-.3156	-.2823	-.2517	-.2821	-.2286
		RV (b)	-.1071***	-.1064***	-.1601***	-.2075***	-.1934***
	V-Ratio	a/e	1.2022	1.0753	.9588	1.0746	.8707
		b/e	**.4081**	**.4052**	**.6099**	**.7906**	**.7369**
-2σ	Expected (e)		-.0933				
	Actual	IV (a)	-.1595	-.1430	-.1020	-.1050	-.1578
		RV (b)	-.0248	.0000***	-.0420	-.0578	-.0449
	V-Ratio	a/e	1.7101	1.5328	1.0934	1.1257	1.6915
		b/e	**.2659**	**.0000**	**.4504**	**.6202**	**.4819**

Note: 1. The lowest shortfall is boldfaced in the column (4)-(8).
2. *, **, *** denotes significance at the 10%, 5% and 1% level, respectively to accept the alternative hypothesis.

Due to the availability of data for OVX, EVZ and GVZ, the sample period in this paper only covered the post-crisis period after June 2008. With this common sample span, this paper could make comparison among the volatility indexes more time-consistently. However, as noted in Table1, VIX and VXN are available since Jan.2, 1990 and Jan.23, 2001, respectively. Since this paper claims a better predictability of VIX and VXN, it will be meaningful to verify whether the superior predictability of VIX and VXN is robust for other periods. For this purpose, this paper conducted an additional test for VIX and VXN over a longer-term sample period using the same methodology.

The outperformance of VIX and VXN was also confirmed by tests for the longer-term sample. Table 8 and 9 report that VIX and VXN performed well in predicting the price interval in both tails. In the right tail, the actual number of violation of both VIX and VXN was less than the expected number of violation and furthermore, even less than that of RV at both intervals. In the left tail, the actual number of violation was higher than RV's but still lower than expected at 1% significance level. Table 10 and 11 present the back-testing results for ES. The results support Table 6 and 7: VIX and VXN outperform both ES of the benchmark and RV in the right tail. One difference is that VIX outperformed ES of the benchmark even in the left tail at 68.3% confidence level. Based upon the results for the longer-term sample, this paper assesses that the outperformance of VIX and VXN are robust and even intensified in the longer-term.

Table 8: Backtesting Results for Price Interval (Right Tail) (Longer Term)

Interval	# of violation		Underlying Index	
			SPX	NDX
			Jan.2, 1990- Mar.31, 2017 (6861 observations)	Jan 23, 2001- Mar.31, 2017 (4072 observations)
+σ	Expected (e)		1089	646
	Actual	IV (a)	504***	444***
		RV (b)	1116	685
	V-Ratio	a/e	**.463**	**.687**
		b/e	1.025	1.061
+2σ	Expected (e)		156	93
	Actual	IV (a)	5***	1***
		RV (b)	44***	47***
	V-Ratio	a/e	**.032**	**.011**
		b/e	.282	.507

Table 9: Backtesting Results for Price Interval (Left Tail) (Longer Term)

Interval	# of violation		Underlying Index	
			SPX	NDX
			Jan.2, 1990- Mar.31, 2017 (6861 observations)	Jan 23, 2001- Mar.31, 2017 (4072 observations)
-σ	Expected (e)		1088	646
	Actual	IV (a)	499***	391***
		RV (b)	388***	262***
	V-Ratio	a/e	.458	.605
		b/e	**.356**	**.406**
-2σ	Expected (e)		156	93
	Actual	IV (a)	66***	59***
		RV (b)	6***	0***
	V-Ratio	a/e	.423	.637
		b/e	**.038**	**.000**

Note: 1. A better interval estimator is boldfaced in the column (4) and (5).
2. *, **, *** denotes significance at the 10%, 5% and 1% level, respectively to accept the alternative hypothesis.

To investigate the reasons why VIX and VXN have been relatively superior in forecasting the right tail of their respective underlying asset, this paper runs multiple regressions of a daily percentage return of volatility index against its lagged index level, positive (r_t^+) and negative percentage return (r_t^-) of its underlying asset price. The regression equation (11) is used to identify the main determinants that affect the return of the volatility index:

$$ln(IV_t/IV_{t-1}) = \alpha + \beta_1 IV_{t-1} + \beta_2 r_t^+ + \beta_3 r_t^- \quad (11)$$

where $r_t^+ = r_t^+ \; if \; r_t^+ > 0,$ $\qquad r_t^- = r_t^- \; if \; r_t^- < 0$

$= 0 \; otherwise \qquad = 0 \; otherwise$

In the equation (11), β_1 captures a persistency in the level of the volatility index. β_2 and β_3 measure an asymmetric relation between asset return and volatility index.

Table 10: Backtesting Results For ES (Right Tail) (Longer Term)

Interval	ES		Underlying Index	
			SPX	NDX
			Jan.2, 1990- Mar.31, 2017 (6861 observations)	Jan 23, 2001- Mar.31, 2017 (4072 observations)
+σ	Expected (e)		1.2626	
	Actual	IV (a)	1.1265***	1.1200***
		RV (b)	1.2020***	1.2251***
	V-Ratio	a/e	**.8922**	**.8870**
		b/e	.9519	.9703
+2σ	Expected (e)		1.0932	
	Actual	IV (a)	1.0457	1.0345
		RV (b)	1.0514*	1.0389**
	V-Ratio	a/e	**.9565**	**.9463**
		b/e	.9618	.9503

Table 11: Backtesting Results for ES (Left Tail) (Longer Term)

Interval	ES		Underlying Index	
			SPX	NDX
			Jan.2, 1990- Mar.31, 2017 (6861 observations)	Jan 23, 2001- Mar.31, 2017 (4072 observations)
-σ	Expected (e)		-.2625	
	Actual	IV (a)	-.2315**	-.2379
		RV (b)	-.1145***	-.1177***
	V-Ratio	a/e	.8819	.9062
		b/e	**.4360**	**.4485**
-2σ	Expected (e)		-0.0933	
	Actual	IV (a)	-.1205	-.1146
		RV (b)	-.0257	.0***
	V-Ratio	a/e	1.2917	1.2292
		b/e	**.2752**	**.0**

Note: 1. A better interval estimator is boldfaced in the column (4) and (5).
2. *, **, *** denotes significance at the 10%, 5% and 1% level, respectively to accept the alternative hypothesis.

In Table 12, β_1 is slightly negative for all IVs at 1% significance level. This result indicates that IVs have a tendency to decline slowly over time. More importantly, it is found that there exists a greater negativeness of β_3 than β_2 for VIX and VXN, which confirms the asymmetric negative relationship between return and IV - i.e., negative return of underlying asset are correlated with positive change in its volatility index, and vice versa, with a greater effect when returns decline/volatility index increases (Hibbert et al. 2008; Han et al. 2012). Also the explanatory power for the return-IV relationship is clearly much higher in VIX and VXN with the Adj-R^2 to be over 0.56.

Table 12. Regression Results: Determinants for IV (June 2008 - March 2017)

	VIX	VXN	OVX	EVZ	GVZ
α	.0151*** (.0027)	.0137*** (.0029)	.0128*** (.0024)	.0106*** (.0029)	.0143*** (.0034)
β_1	-.001*** (.0002)	-.0009*** (.0001)	-.0008*** (.0001)	-.002*** (.0003)	-.0019*** (.0002)
β_2	-2.956*** (.2962)	-2.4959*** (.242)	.2736** (.1299)	.8327* (.501/)	2.6135*** (.2525)
β_3	-5.0457*** (.4276)	-4.1545*** (.241)	-1.8987*** (.1185)	-4.1035*** (.3471)	-3.3743*** (.3164)
$Adj-R^2$	.567	.5633	.2588	.1141	.2035

Note: *, **, *** denotes significance at the 10%, 5% and 1% level, respectively to accept the alternative hypothesis.

The superior forecasting power of VIX and VXN for the right tail will be related to this asymmetric negative return-IV relation in the stock market, which is explained by two hypothesis. The 'leverage effect' signifies that a fall in the equity value increases the leverage of firm and thus its equity risk. The 'volatility feedback hypothesis' (French et al., 1987; Campbell & Hentschel 1992) states that volatility feedback amplifies large negative returns and dampens large positive returns and thus it increases the potential for large crashes. That is, due to the leverage effect, the negative daily return of the stock market increases its volatility index and then the gap between its upper and lower bound for the predicted price interval is wider, and vice versa. In this circumstance, the volatility feedback intensifies the negative return and softens the positive return, which decreases the potential for the NDEV's deviation above the upper bound. By contrast, the assets in the non-stock markets are not subject to the leverage effect so that OVX, EVZ and GVZ are not revealing the asymmetric relationship: i.e., both up and down in the daily return of its underlying asset increase the IVs. This result supports Padungsaksawasdi and Daigler (2014) who examined the return-implied volatility relation for euro, gold and oil and found that gold possesses a positive contemporaneous return coefficient and that the euro and gold are not asymmetric.

In sum, this paper rejectsthe null hypothesis that IV cannot be used in predicting the price interval at least for VIX and VXN. VIX and VXN also achieved lower ES than that of RV and benchmark in the upper bound. The test results for USO, FXE and GLD are mixed: none of them statistically have achieved better than RV and benchmark in the lower bound.

5. Conclusion

Volatility indexes have been popularly used to predict a point forecast for the volatility of financial products. However, it has been criticised that the point forecast has a limited value due to its only one side description of possible outcome. Correspondingly, a more accurate prediction for the tail distributions represented by VaR and the expected shortfall (ES) is of particular importance for ETF portfolio investors. In this context, this paper investigated whether the five CBOE volatility indexes (IV) can be used in estimating VaR. Also to convince the usage of IV for VaR, this paper additionally tested whether ES beyond the price interval forecasted by each IV is lower than ES of the ex-post volatility (RV) and its normal distribution benchmark.

Several prior studies reported that the point forecasting power of volatility index or implied volatility were controversial.[10] However, with a focus on the tail distribution of returns, this paper found several interesting results on the forecasting power of volatility index for the price interval. The findings of this paper are highlighted as follows: first, VIX and VXN exhibited a good forecasting power for the price interval of upper bound (i.e., right tail). VIX and VXN were even superior to the ex-post volatility (RV) of its corresponding underlying asset. VIX and VXN also achieved lower ES than that of RV and benchmark in the upper bound. Second, VIX and VXN proved a good forecasting power for the price interval of the lower bound (i.e., left tail). Third, the forecasting power of VIX and VXN is also robust in a longer-term sample period. The outperformance of VIX and VXN for the right tail predictability can be explained by the asymmetric negative return-IV relation in the stock market, which is explained by the 'leverage effect' and the 'volatility feedback hypotheses.

In contrast, the forecasting power of OVX, EVZ and GVZ were inconsistent and inferior to VIX and VXN. Nonetheless, it is uncertain whether the inconsistent or inferior performance is attributable to some idiosyncratic time-varying market conditions during the sample period or the less efficiency in the market for its underlying asset.

To the best of the author's knowledge, this paper is the first comparative analysis to examine the forecasting power of the CBOE's five volatility indexes for the price interval and the expected shortfall. Using a practical and easily implementable method, this paper showed that VIX and VXN can be used to predict the price interval or VaR. The findings of this paper is expected to provide ETF mangers and investors for US stock markets with useful information in predicting the one-month price range of their portfolio.

However, the results of this paper must be accepted with some caveats because historical simulation is sensitive to sample periods. In particular, the analysis for OVX, EVZ and GVZ was confined to the post-crisis period after June 2008 because of the data availability. In addition, due to the space limit, this paper could not go into a deeper analysis for each volatility index and the interactions among the volatility indexes. Complementing these limitations will be potential sources of topics for further research.

Endnotes

[1] See Patton & Sheppard (2009) and Patton (2011) for the detail of the limitations.
[2] See The CBOE (2015) for the VIX algorithm. OVX, EVZ and GVZ can be practically used to predict the volatility of the WTI crude oil price, the USD/EUR exchange rate and the Gold price, respectively. However, to avoid the bias due to different time zone between a volatility index and its underlying asset, this paper used the data of its underlying ETF (i.e., USO, FXE and GLD) for each volatility index. The substitution of each underlying EFT by its physical asset will be also acceptable for the analysis because the correlations of each substitute's volatility are considerably high for the sample period (i.e., the WTI oil price vs. USO: 0.963, the USD/EUR exchange vs. FXE: 0.944 and the Gold price vs. GLD: 0.924).
[3] For example, see Haugom et al. (2014) for OVX and Luo et al. (2016) for GVZ.
[4] See Diebold & Lopez (1996) and Gneiting (2011) for a theoretical framework of point forecasts.
[5] See Campbell (2007) and Abad et al. (2014) for the review on the VaR methodologies and backtesting procedures.
[6] According to Wong (2008), a coherent risk measure is monotonous (systematic lower returns imply greater risk), sub-additive (sum of individual risks cannot be smaller than risk of sums), positively homogenous (larger positions are associated with greater risk) and translation invariant (investing in risk-free asset will lower the risk of portfolio).
[7] The data for the volatility indexes and underlying indexes are retrieved from CBOE microsite http://www.cboe.com/products/and https://finance.yahoo.com/. The data for WTI Crude oil, EUR and Gold are obtained from https://fred.stlouisfed.org/.

[8]According to The CBOE (2015), $VIX = 100 \times \sqrt{\left\{T_1 \sigma_1{}^2 \left[\frac{N_{T_2} - N_{30}}{N_{T_2} - N_{T_1}}\right] + T_2 \sigma_2{}^2 \left[\frac{N_{30} - N_{T_1}}{N_{T_2} - N_{T_1}}\right]\right\} \times \frac{N_{365}}{N_{30}}}$

where $\sigma_1{}^2$= volatility of near-term put and call options for SPX, $\sigma_2{}^2$= volatility of next-term put and call options for SPX,T_1= the time weight of the near-term options, T_2= the time weight of the next-term options, N_{T1} = number of minutes to settlement of the near-term options, N_{T2}= number of minutes to settlement of the next-term options, N_{30}= number of minutes in a 30 days, N_{365}= number of minutes in a 365-day year. This paper uses 21 trading days as equivalent to 30 calendar days because the average of annual trading days of the NYSE and NASDAQ is approximately 252. Therefore, when the realized volatility is calculated, it needs to multiply the daily average of squared returns by the square root of 365to maintain the consistency with the above formula (i.e., $\sqrt{\frac{N_{365}}{N_{30}}}$) that the CBOE calculates the VIX index. See Corrado & Miller (2005); Carr & Wu (2006) for the papers using the similar formula.

[9] While Kupiec (1995), McNeil & Frey (2000) and Mancini & Trojani (2011) used the two-sided binomial test, this paper performed a one-sided binomial test separately for each tail.

[10] Based upon a survey on 66 papers that provided comparisons among estimated volatilities, Poon & Granger (2005) found that implied volatility is more accurate than time-series models because the market option price fully incorporates current information and expectation for future volatility. Yet, many other papers presented a question about the domination of the implied volatility (or volatility index) over other volatility estimators. For example, see Figlewski (1997), Agnolucci (2009), Bentes (2015) and Jung (2016).

References

Abad, P, Benito, S & López, C 2014, ‘A comprehensive review of value at risk methodologies’, *The Spanish Review of Financial Economics*, vol.12,no.1, pp.15-32.

Acerbi, C & Tasche, D 2002a, ‘Expected shortfall: A natural coherent alternative to value at risk’, *Economic Notes*, vol.31, no.2, pp.379–388.

Acerbi, C & Tasche, D 2002b, ‘On the coherence of expected shortfall’, *Journal of Banking and Finance*, vol.26, no.7, pp.1487–1503.

Acerbi, C & Szekely, B 2014, ‘Back-testing expected shortfall’, *Risk*, vol.27, no.11, pp.76–81.

Agnolucci, P 2009, ‘Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models’, *Energy Economics*, vol.31, no.2, pp.316–321.

Amisano, G & Giacomini, R 2007, ‘Comparing density forecasts via weighted likelihood ratio tests’, *Journal of Business and Economic Statistics*, vol.25, no.2, pp.177–190.

Andersen, TG & Bollerslev, T 1998, ‘Answering the skeptics: yes standard volatility models do provide accurate forecasts’, *International Economic Review,* vol.39, no.4, pp. 885-905.

Artzner, P, Delbaen, F, Eber, JM & Heath, D 1997, ‘Thinking coherently’, *Risk*, vol.10, no.11, pp.68-71.

Artzner, P, Delbaen, F, Eber, JM & Heath, D 1999, ‘Coherent measures of risk’, *Mathematical Finance*, vol.9, no.3, pp.203–228.

Bentes, S 2015, ‘A comparative analysis of the predictive power of implied volatility indices and GARCH forecasted volatility’, *Physica A*, vol.424,pp.105-112.

Berkowitz, J 2001, ‘Testing density forecasts, with applications to risk management’, *Journal of Business and Economic Statistics,* vol.19, no.4, pp.465–74.

Campbell, S 2007, ‘A review of backtesting and backtesting procedures’, *Journal of Risk*, vol.9, no.2, pp.1–17.

Campbell, J & Hentschel, L 1992, ‘No news is good news: an asymmetric model of changing volatility in stock returns’, *Journal of Financial Economics*, vol.31, no.3, pp.281- 318.
Carr, P & Wu, L 2006, ‘A tale of two indices’, *Journal of Derivatives,* vol.13, no.3, pp.13–29.
Cesarone, F & Colucci, S 2016, ‘A quick tool to forecast VaR using implied and realized volatilities, working paper, SSRN: http://ssrn.com/abstract=2714443 orhttp://dx.doi.org/10.2139/ssrn.2714443
Chatfield, C 1993, ‘Calculating interval forecasts, ‘*Journal of Business and Economics Statistics*, vol.11, no.2, pp.121-135.
Chen, Q & Gerlach, RH 2013, ‘The two-sided Weibull distribution and forecasting financial tails’, *International Journal of Forecasting*, vol.29, no,.4, pp.527–540.
Christoffersen, PF 1998, ‘Evaluating interval forecasts’, *International Economic Review*, vol.39, no.4, pp. 841–862.
Corrado, CJ & Miller, TW 2005, ‘The forecast quality of CBOE implied volatility indexes’, *Journal of Futures Markets*, vol.25, no.4, pp.339–373.
Costanzino, N & Curran, M 2015, ‘Backtesting general spectral risk measures with application to expected shortfall’, *Journal of Risk Model Validation*, March, pp.21-31.
Diebold, FX &Lopez, JA 1996, ‘Forecast evaluation and combination’, In Maddala GS, Rao CR (eds), *Handbook of Statistics*, North-Holland: Amsterdam, pp. 241–268.
Diebold, FX, Gunther, TA & Tay, AS 1998, ‘Evaluating density forecasts with applications to financial risk management’, *International Economic Review*, vol.39, no.4, pp.863–883.
Emmer, S, Kratz, M & Tasche, D 2015, ‘What is the best risk measure in practice? A comparison of standard measures’, *Journal of Risk*, vol.18, no.2, pp.31-60.
Figlewski, S1997, ‘Forecasting volatility’. *Financial Markets, Institutions & Instruments,* vol. 6, no.1, pp.1-88.
French, K, Schwert, G & Stambaugh, R 1987, ‘Expected stock returns and volatility’, *Journal of Financial Economics*, vol.19, no.1, pp.3-29.
Garman, M & Klass, M 1980, ‘On the estimation of security price volatilities from historical data’, *Journal of Business,* vol.53, no.1, pp.67–78.
Giot, P 2005, ‘Implied volatility indexes and daily value at risk models’, *Journal of Derivatives*, vol.12, pp.54-64.
Gneiting, T 2011, ‘Making and evaluating point forecasts’, *Journal of the American Statistical Association*, vol.106, pp.746–762.
Han, Q, Guo, B, Ryu, D & Webb, RI 2012, ‘Asymmetric and negative return- volatility relationship: The case of the VKOSPI’, *Investment Analysts Journal*, vol.41, no.76, pp.69–78.
Haugom, E, Langeland, H, Molnár, P & Westgaard, S 2014,‘Forecasting volatility of the US oil market’, *Journal of Banking & Finance*, vol.47, pp.1–14.
Hibbert, A, Daigler, R & Dupoyet, B 2008, ‘A behavioral explanation for the negative asymmetric return-volatility relation’, *Journal of Banking & Finance*, vol.32, no.10, pp.2254–2266.
Jung, Y 2016, ‘Relative performance of VIXC vs. GARCH in predicting realized volatility changes’, *Investment Analyst Journal*, vol.45, no.sup1, pp.s1-s16.
Kerkhof, J & Melenberg, B 2004, ‘Backtesting for risk-based regulatory capital’, *Journal of Banking & Finance*, vol.28, no.8, pp.1845–1865.

Kim, JS& Ryu, D 2015, 'Are the KOSPI 200 implied volatilities useful in value-at-risk models?', *Emerging Markets Review*, vol.22, pp.43-64.

Koopman, SJ, Jungbacker, B & Hol, E 2005, 'Forecasting daily variability of the S&P 100 stock index using historical, realized and implied volatility measurements', *Journal of Empirical Finance*, vol.12, no.3, pp.445–475.

Kristoufek, L 2014, 'Leverage effect in energy futures', *Energy Economics*, vol.45, pp.1–9.

Kupiec, PH 1995, 'Techniques for verifying the accuracy of risk measurement models', *Journal of derivatives*, vol.3, no.2, pp.73-84.

Luo, X, Qin, S & Ye, Z 2016, 'The information content of implied volatility and jumps in forecasting volatility: evidence from the Shanghai gold futures market', *Finance Research Letters*, vol.19, pp.105–111.

Mancini, L & Trojani, F 2011, 'Robust value at risk prediction', *Journal of Financial Econometrics*, vol.9, no.2, pp.281–313.

Martens, M2002, 'Measuring and forecasting S&P 500 index-futures volatility using high-frequency data', *Journal of Futures Markets*, vol.22, no.6, pp.497 – 518.

Martens, M & Zein, J 2004, 'Predicting financial volatility: high-frequency time-series forecasts vis- à-vis implied volatility', *Journal of Futures Markets*, vol.24, pp.1005–1028.

McNeil, AJ & Frey, R 2000, 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance*, vol.7, no.3-4, pp.270-300.

Morgan, JP 1996, *Riskmetrics-technical document. Tech. rep*., New York: Morgan Guaranty Trust Company of New York, 4th ed.

Padungsaksawasdi, C & Daigler, RT 2014, 'The return-implied volatility relation for commodity ETFs', *Journal of Futures Markets*, vol.34, no.3, pp.261–281.

Patton, AJ & Sheppard, K 2009, 'Evaluating volatility and correlation forecasts', In Andersen, TG, Davis, RA, Kreiss, JP & Mikosch, T (eds.), *The Handbook of Financial Time Series*. Springer Verlag, pp.801-838.

Patton, AJ 2011, 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics*, vol.160, no.1, pp.246–256.

Pedersen, TQ 2015, 'Predictable return distributions', *Journal of Forecasting*, vol.34, pp.114–132.

Poon, S & Granger, C 2005, 'Practical Issues in Forecasting Volatility', *Financial Analyst Journal,* vol.61, no.1, pp.45-56

Rubia, A & Sanchis-Marco, L 2013, 'On downside risk predictability through liquidity and trading activity: A dynamic quantile approach', *International Journal of Forecasting*, vol.29, no.1, pp.202–219.

Tay, AS & Wallis, KF 2000, 'Density forecasting: A survey', *Journal of Forecasting*, vol.19, no.4, pp.235 – 254.

Taylor, SJ & Xu, X 1997, 'The incremental volatility information in one million Foreign Exchange Quotations', *Journal of Empirical Finance*, vol.4, pp.317–340.

The CBOE 2015, *White Paper: The CBOE Volatility Index – VIX*.

Wong, WK 2008, 'Backtesting trading risk of commercial banks using expected shortfall', *Journal of Banking & Finance*, vol.32, no.7, pp.1404–1415.

Yamai, Y & Yoshiba, T 2002, 'Comparative analyses of expected shortfall and VaR: Their estimation error, decomposition, and optimization', *Monetary and Economic Studies*, vol.20, no.1, pp.87–121.

Yamai, Y & Yoshiba, T 2005, 'Value-at-risk versus expected shortfall: A practical perspective', *Journal of Banking & Finance,* vol.29, no.4, pp.997–1015.

Yang, Z & Zhang, Q 2000, 'Drift-independent volatility estimation based on high, low, open, and close prices', *Journal of Business*, vol.73, no.3, pp.477–491.

Appendix

Table A1. A Brief Overview of the Five CBOE Volatility Indexes (IV)

Volatility Index	Underlying Asset	IV Data Available Since
Volatility Index (VIX)	S&P500 (SPX)	Jan.2, 1990
Nasdaq-100 Volatility Index (VXN)	Nasdaq100 (NDX)	Jan.23, 2001
Crude Oil ETF Volatility Index (OVX)	United States Oil Fund, LP (USO)	May 10, 2007
EuroCurrency ETF Volatility Index (EVZ)	CurrencyShares Euro Trust (FXE)	Nov.1, 2007
Gold ETF Volatility Index (GVZ)	SPDR Gold Shares (GLD)	June 3, 2008