The VIX index: forecasting power and performance in a risk management framework

DOI:10.12831/85433

The VIX Index: Forecasting Power and Perfomance in a Risk Management Framework

Alessio Bongiovanni[1], Paola De Vincentiis[2], Eleonora Isaia[3]

Abstract

This paper explores the information content and forecasting power of the VIX Index computed by CBOE (Chicago Board Options Exchange) on two different levels. The analysis is organised around two research questions. The first question is aimed at understanding whether the Volatility Index (VIX), due to its forward-looking nature, forecasts the future realised volatility better than other estimation techniques that are based on historical data. This part of the analysis is in line with a rich stream of literature on the topic, and our contribution intends to at test whether the empirical results obtained by other researchers hold true in the years of hightened volatility following the Lehman Brothers collapse. The second research question aims to evaluate the perfomance of the VIX

[1] Department of Management, University of Torino, mail: alessio.bongiovanni@unito.it

[2] Department of Management, University of Torino, 218 mail: paola.devincentiis@unito.it

[3] Department of Management, University of Torino, 218 bis, mail: eleonora.isaia@unito.it

within a risk management framework, exploring an aspect that has been scarcely analysed in the literature and that has produced relatively contradictory results. In particular, we use the VIX alongside other volality measures to compute the Value-at-Risk (VaR) metric for a hypothetical portfolio replicating the Standard & Poor's 500 Index. The various measures of maximum potential loss are then backtested against actual returns and compared in order to understand which one is more effective.

Results show that the VIX index possesses a strong information content, even if it is an upward biased forecast of realised performance. When used to compute VaR however, the measures based on VIX are less effective than others using different volatility estimations, especially during periods of higher volatility.

Keywords: VIX, historical volatility, GARCH models, forecast ability, information content
JEL Classification: G14, G17

1 Introduction

Estimating volatility is one of the main goals of academics and practitioners in the financial field. Forecasts of future price variability are needed to make funding or investment decisions, to value financial instruments, and to measure the risk of financial asset portfolios. Not surprisingly, a vast empirical and theoretical literature has focused on this topic proposing new methods for estimating volatility or comparing the effectiveness of techniques already in use. In particular, our work belongs to the stream of literature that explores the merits of implied volatility (IV) measures, i.e. volatility measures derived from option prices. From a theoretical point of view, these measures could be superior to other types of estimates because they reflect market expectations instead of being derived from a statistical model or from historical returns. In fact, IV is often indicated as a forward-looking measure.

The analysis conducted and described in this paper revolves around two research questions. On the one hand, we want to judge comparatively the forecasting power of the VIX, i.e. its ability to approximate future realised volatility in relation to other estimation techniques. There is quite a rich empirical literature on this topic and our incremental contribution is twofold. First, we want to determine if the ability to approximate future realised volatility improved when options contracts on VIX began to be traded in 2006. Second, we aim to understand whether the VIX performance is better in periods of high or low market volatility. The period analysed – which includes the turbulent years following the Lehman Brothers collapse – represents a perfect "natural experiment" to this end.

The second research question aims to evaluate the performance of the VIX in a risk management framework, exploring an aspect that has been scarcely analysed in the literature and with quite contradictory results. In particular, we compute Value-at-Risk (VaR) for a hypothetical portfolio that is a replication of the Standard & Poor's 500 Index both using VIX and other alternative volatility estimates. We then backtest and compare the various risk measures that were computed in order to understand which one performs better. This exercise is conducted for two sub-periods that are characterised by high and low market volatility.

In summation, our results confirm the strong informative value of VIX and its superior forecasting ability, once a problem of outliers is dealt with. The analysis shows however, that the VIX is an upward biased estimate of realised volatility. When used to compute VaR however, the measures based on the VIX are less effective than others using different volatility estimations, especially in periods of higher volatility.

As mentioned above the literature concerning volatility measurement is rich and extensive. One stream of literature – in line with our first research question – compares various volatility-forecasting methods by pitting one against the other. Typically, the expected volatility estimated through different alternative methods is used as an independent variable to explain realised volatility, i.e. the dependent variable. The information content and forecasting power of the expected volatility measure are judged by examining the significance of the beta coefficient and by testing the null hypothesis that the coefficient is equal to 1 and the intercept is equal to zero. The relative forecasting power of different volatility measures is analysed by including them concurrently in a regression, and by comparing the coefficients of the various independent variables.

Poon and Granger (2005) examined 93 studies that were structured in this way and published over a 20-year period. Their overall conclusion is that option-implied volatility most frequently provides better forecasts than time-series models. Jorion (1995) is among the most influential empirical studies dealing with option-implied volatility. Focusing on the currency market, he finds that implied volatility outperforms a statistical time-series, even when they are given the advantage of *ex-post* parameter estimates. However, IV appears to be a biased volatility forecast. Similarly, Fleming (1998), Ederington and Guan (2002), Szakmary et al. (2003), Corrado and Miller (2005) find that IV dominates historical volatility despite being an upward biased forecast. Shu and Zhang (2003) reach the same conclusion using four different measures of realised volatility that are characterised by increasing complexity. Day and Lewis (1992) find that implied volatilities derived from options on the S&P 100 index contain incremental information when added as an exogenous variable to GARCH and EGARCH models, but they are unable to draw precise conclusions as to the relative predictive power of GARCH forecasts and implied volatility to *ex post* volatility.

Canina and Figlewski (1993) sharply refute the papers mentioned above. Indeed, they find that implied volatility derived from S&P 100 index options has no correlation with future volatility at all. However a few years later, Christensen and Prahbala (1998) strongly criticised the methodology of this study, attributing its peculiar results to a problem of overlapping data that was not adequately managed. By solving the issue, the authors confirm that implied volatility outperforms historical volatility in forecasting future volatility, and this provided stronger evidence when compared to previous studies. Further confutations are made by Becker et al. (2007) who find that the VIX index does not contain incremental information when compared to a combination of model-based volatility forecasts. As in the study conducted by Canina and Figlewski (1993), this empirical study presents a problem of overlapping observations. Moreover, they do not directly compare VIX forecasts against any single model-based forecast but against quite a complicated combination that would be difficult to use in daily practice. Thus, the contribution is merely theoretical.

The most recent contributions in the literature focus on comparing the performance of different models across different asset classes, different financial markets, and in different market conditions. Kourtis et al. (2016) compare the forecasting power of implied and GARCH volatility at an international level, taking into consideration 13 equity indices from 10 countries. They find a very similar ranking of the models analysed in the different markets. In particular, the implied volatility corrected for the risk premium is the best monthly forecast, whereas the Heterogeneous Autoregressive model is superior with regard to the daily time horizon.

Browless et al. (2011) compare a set of models belonging to the ARCH family on a wide array of assets with the aim of comparing not only their forecasting power, but also their ability to cope with a period of financial market crisis such as the 2008 turmoil. Surprisingly, they find a ranking that is insensitive to market conditions at the daily horizon. On the contrary, long-run forecasts are negatively affected by a surge in market volatility.

Charoenwong et al. (2009), focusing on the foreign exchange market, compare the predictive power of implied volatility derived from exchange-traded and over-the-counter options, concluding that the latter are superior. Furthermore, they confirm the greater information content of implied volatility compared to time-series based estimates.

Moving to our second research question, a few papers have previously tested the effectiveness of different volatility forecasts in a risk management framework. In this case, the volatility forecast is not directly compared to realised volatility but is used as an input to estimate the maximum potential loss on a given asset or portfolio with a determined confidence level. The effectiveness of the measure of risk is then backtested against actual losses. In these empirical works, the focus is not on the capacity to predict the level of volatility exactly, but on the ability to adequately capture the tails of the distribution, and as such, the extreme values. Christoffersen et al. (2001) find no evidence that VaR estimated using implied volatility is superior to the same indicator based on GARCH or historical volatilities. Chong (2004) finds that implied volatility is not effective in estimating VaR because it tends to overestimate volatility in periods of stability and underestimate risk when the market is more volatile. Conversely, Giot (2005) finds that IV performs quite well as an input to VaR measurement, even in challenging market conditions characterised by bear prices and high volatility. More recently in a study focused on the Korean market, Kim and Ryu (2015) document a rather poor performance in VaR estimates for the equivalent of the VIX on the KOSPI index, when compared to GARCH-based volatilities or to implied volatilities directly derived from OTM or ATM options. This poor performance is particularly evident during and after the sub-prime crisis when the models based on ATM implied that volatilities outperform alternative estimation methods.

Proceeding this brief review of the literature on the topic, in the following sections we describe the methodology adopted by this study and the features of the data sample (section 2), present the results of our empirical investigation (section 3) and draw conclusions (section 4).

2 Methodology and sample

The first research question of our paper explores the information content and the predictive power of the VIX index. We investigate the relation between implied and realised volatility, and assess whether the VIX index is a better predictor of future volatility compared to historical and GARCH-based volatility measurements.

In the analyses, we use daily closing prices calculated directly by the CBOE, which represent the implied volatilities of the S&P 500 index over the next 30-day period (22 trading days). Our analyses examine a 20-year period from January 1995 to December 2014, divided into two sub-periods before and after March 2006, which represent the introduction of options on the VIX index. By comparing the two sub-periods, we can judge whether the information content of the VIX increased after becoming a negotiable asset.

We initially run univariate regressions considering the realised volatility as a dependent variable, and the VIX index and the other methodologies that are based on historical data, as independent variables. For brevity these analyses are not fully reported, but their results are essentially in line with the successive analysis and with the majority of previous studies.

We then include both implied and historical volatilities in a multivariate regression analysis following the main stream of the literature on this topic. The following equations were estimated:

$$RV_t = \alpha + \beta \, Vix_{t-1} + \beta \, SMA_{t-1} \qquad (1)$$

$$RV_t = \alpha + \beta \, Vix_{t-1} + \beta \, EWMA_{t-1} \qquad (2)$$

$$RV_t = \alpha + \beta \, Vix_{t-1} + \beta \, GARCH_{t-1} \qquad (3)$$

where, VIX = Volatility index computed by CBOE; SMA = simple moving average, computed on the 22 most recent daily returns of the S&P 500 Index; EMWA = exponential moving average computed on the 22 most recent daily returns using as decay factor calibrated on the historical return on the three previous years; GARCH: volatility computed with a GARCH

(1;1) model with parameters calibrated with 5 years of historical returns for the model applied in the entire period and calibrated with 3 years for the ones used in the two sub-periods.

As part of the empirical analysis, two problems had to be dealt with: overlapping data (Canina & Figlesky, 1993; Christensen & Prabhala, 1998) and possible errors in the realised volatility measurement. To address the first issue, we considered the VIX price of the day following the measurement of the realised volatility for each period, which is calculated again after 22 trading days.

To manage the second problem, we tested four different measurements of realised volatility that are gradually more accurate,, namely the standard deviation, the Parkinson extreme value estimator (1980), the Roger and Satchel estimator (1991), and the Yang and Zhang estimator (2000). We ran a regression analysis for each of the different measurements of realised volatility, considering each in turn as a dependent variable.

Though the majority of the studies on this topic do not deal with the multi-collinearity problem that might arise when the VIX index and a measure of historical volatility enter the same model, we prefer to face this issue by computing and evaluating the Variance Inflation Factors. In fact, a potential imperfect collinearity between these two variables cannot be excluded *a priori*. In this regard, it should be mentioned that few abnormal observations — registered at the heart of the financial crisis, from September 2008 to April 2009, and identified both with the leverage measure and Cook's distance — have been excluded from the regressions in order to reduce the multi-collinearity effect.

Moving to the second research question, we tested the feasible use of the VIX to estimate the maximum potential loss in a VaR model and its supposed greater ability with respect to the GARCH, SMA and EWMA historical measurements. We thus computed the VaR on a daily basis, of a hypothetical portfolio replicating Standard & Poor's 500 Index, considering the four volatility estimators one after the other. Backtesting procedures were used to compare the different models. In particular, we contrasted two periods: January 2008 to December 2009 (504 observations), named the 'high-volatility period', and March 2013 to February 2015 (499 observations), named the 'low-volatility period'. The former is the length of time during which the VIX index reached its peak, whereas the latter is the time frame when the market volatility was extremely low and rarely above 20%. We then assessed the accuracy of the models by carefully analysing the exception rate, using Kupiec's unconditional coverage test (1995), Christoffersen's conditional coverage test (1998), and Lopez's loss function test (1999).

Table 1 refers to the first research question and provides some descriptive statistics for the volatility estimation methodologies, comparing the VIX index to the traditional methods that are based on historical data. Despite the crisis market phase experienced in 2008–2009, mean and median values do not present significant differences among the periods analysed, remaining quite similar even when the entire sample is split into the two sub-samples – before and after the introduction of VIX option trading in 2006.

Indeed, the only elements that prove the stressed conditions characterising the second sub-period (2006–2014) are the larger standard deviation of each estimation method, the maximum values, which are considerably higher, and the higher root mean square errors (RMSE) that indicate more difficulty in predicting the realised volatility compared to the previous period.

The higher mean and median values registered by the VIX Index in the entire timespan seem to suggest an upward bias, which might incorporate a greater weight given by investors to the occurrence of low frequency/high impact losses. It is also worth noting – at this preliminary descriptive level – the higher RMSE associated with the VIX in the second sub-period (2006–2014) that contains the subprime crisis, and is consequently characterised by higher peak values of volatility.

Table 1. Descriptive statistics for the entire period 01/1995–12/2014 and for the two sub-periods 01/1995–02/2006 and 03/2006–12/2014

Volatility estimators for the period 01/1995–12/2014						
	Mean	Median	Minimum	Maximum	Standard deviation	Mean square
VIX	20.54%	19.61%	10.05%	80.06%	8.41%	0.5421%
GARCH	16.24%	14.08%	7.72%	58.65%	7.68%	0.5076%
SMA	16.56%	14.48%	5.39%	80.76%	9.74%	0.5143%
EWMA	16.67%	14.52%	6.09%	74.43%	9.46%	0.4731%
Volatility estimators for the period 01/1995–02/2006						
	Mean	Median	Minimum	Maximum	Standard deviation	Mean square
VIX	20.29%	20.18%	10.77%	37.52%	6.31%	0.4353%
GARCH	11.66%	10.95%	8.59%	21.75%	2.78%	0.5562%
SMA	15.94%	14.64%	5.84%	44.92%	7.35%	0.4055%
EWMA	16.16%	15.02%	6.14%	40.27%	7.17%	0.3808%
Volatility estimators for the period 03/2006–12/2014						
	Mean	Median	Minimum	Maximum	Standard deviation	Mean square
VIX	20.76%	17.66%	10.05%	80.06%	10.51%	0.6935%
GARCH	16.51%	13.50%	8.26%	61.28%	9.54%	0.6039%
SMA	17.33%	14.24%	5.39%	80.76%	12.06%	0.6981%
EWMA	17.30%	14.27%	6.40%	74.64%	11.67%	0.6036%

For the methodologies based on historical data, the volatility is computed using daily observations and expressed in annualised terms.

Table 2 refers to the second research question and presents the descriptive statistics of the different volatility estimators used as inputs in a VaR model. The differences between the high and low-volatility periods are evident; the comparison between averages, the larger the spread from minimum to maximum – essentially due to some abnormal observations – and the higher the standard deviation in the high-volatility period immediately point out the significantly more volatile phase experienced by the market in the years 2008–2009. Consistently with the data of Table 1, the average VIX level once again indicates its tendency to provide a volatility estimate that is higher than other methods, suggesting that its use as a market risk parameter might lead to a more conservative estimate.

Table 2. Descriptive statistics for the periods on which the VaR models are backtested

Volatility estimators for the high-volatility period 2008–2009					
	Mean	Median	Minimum	Maximum	Standard deviation
SMA	30.73%	23.02%	10.08%	89.95%	18.21%
EWMA	30.25%	22.61%	10.84%	83.55%	17.36%
GARCH	26.97%	20.54%	11.70%	71.14%	14.33%
VIX	32.09%	26.01%	16.30%	80.86%	13.25%
Volatility estimators for the low-volatility period 2013–2015					
	Mean	Median	Minimum	Maximum	Standard deviation
SMA	11.40%	10.74%	5.45%	18.94%	3.26%
EWMA	11.18%	10.56%	6.27%	17.99%	2.82%
GARCH	12.84%	11.99%	8.95%	22.50%	2.75%
VIX	14.50%	13.89%	10.32%	26.25%	2.48%

3 Results

In order to present our findings in the clearest possible way, this section is organised into sub-sections. The sub-sections 3.1 and 3.2 concern the first research question; we initially run a bivariate regression in order to compare the information content of both VIX and the historical methods by entering them as independent variables within the same regression. This part of the analysis is quite standard in the related literature. We then provide some innovations that build upon the previous studies by dealing with collinearity problems. The sub-section 3.3 refers to our second research goal that aims to evaluate the performance of the VIX in a risk management framework, and we report the performance test of the various estimation methods as input into VaR models.

3.1 First research question: comparison between the predictive power of the various estimation methods

As the first step in our analysis, we pitted the VIX against historical and GARCH-based volatility to test its relative predictive ability. Tables 3 and 4 present the results of this analysis. In particular, they only report the results obtained using EWMA, however an unreported robustness check made by substituting EWMA with SMA confirms the evidence. Focusing on the entire period, the values of the VIX_{t-1} coefficients range from 0.3739 to

0.9113, are higher than those exhibited by the methods that utilise historical data, indicating a better forecasting ability for volatility derived from option prices. This evidence is confirmed in all regressions, regardless of the method adopted to measure the realised volatility.

When we split the sample using March 2006 as the dividing point, the evidence is strongly confirmed in the first sub-period 01/1995–02/2006 where the higher forecasting ability of the VIX can be seen once again. This sub-period differs from the entire timespan only for the slope coefficients of the historical estimation techniques that are not statistically different from zero, confirming the superiority of the VIX that even incorporates the information of EWMA and GARCH-based volatility. The analysis of the second sub-period 03/2006–12/2014, provides evidence of the weak forecasting performance for the VIX index instead, especially when compared to the EWMA. Surprisingly, the coefficients of EMWA are higher in all the measurement methods that we considered, and VIX coefficients are not significantly different from zero. Thus put simply, the information content of the implied volatility is subsumed by the historical volatility. The GARCH estimates have predictive power as well. Nevertheless, we cannot identify a volatility prediction method that is clearly superior to all the others. In fact, order relations are variable and depend on the measuring techniques analysed. Furthermore, the differences between coefficients are not large enough to say which one presents the better performance.

Table 3. Regression models for the different measures of realised volatility, assuming the VIX level as an independent variable, and the historical volatility computed by the exponential weighted moving average (EWMA)

Dependent variables for the period 01/1995–12/2014				
	$\sigma_{Dev.std}$	σ_{Park}	$\sigma_{R\&S}$	$\sigma_{Y\&Z}$
Intercept	-0.01323	0.00722	0.01496*	0.009609
	(0.0116)	(0.0092)	(0.0086)	(0.0089)
VIX_{t-1}	0.6876**	0.4561**	0.3739**	0.4351**
	(0.1157)	(0.0915)	(0.0857)	(0.0886)
$EWMA_{t-1}$	0.2269**	0.2099**	0.2444**	0.2401**
	(0.1030)	(0.0815)	(0.0764)	(0.0789)
N	227	227	227	227
Adjusted R^2	0.62	0.5832	0.5822	0.6071
F(3, 224)	34.07	182.74	235.27	179.58
Dependent variables for the period 01/1995–02/2006				
Intercept	-0.01757	-0.005486	0.002497	-0.001501
	(0.0151)	(0.0121)	(0.0112)	(0.0116)
VIX_{t-1}	0.8427**	0.6169**	0.5330**	0.5897**
	(0.1246)	(0.0992)	(0.0924)	(0.0953)
$EWMA_{t-1}$	0.03215	0.08877	0.1323	0.1172
	(0.1094)	(0.0871)	(0.0811)	(0.0837)
N	126	126	126	126
Adjusted R^2	0.5651	0.5706	0.5772	0.5908
F(3; 123)	36.21	144.39	180	144.2
Dependent variables for the period 03/2006–12/2014				
Intercept	0.01080	0.02042	0.02826**	0.02287
	(0.01842)	(0.01458)	(0.01367)	(0.01417)
VIX_{t-1}	0.3649*	0.2231	0.1291	0.1988
	(0.2135)	(0.1690)	(0.1584)	(0.1642)

$EWMA_{t-1}$	0.5011**	0.4096**	0.4472**	0.4427**
	(0.1921)	(0.1521)	(0.1426)	(0.1478)
N	101	101	101	101
Adjusted R^2	0.6273	0.5925	0.5865	0.6158
F(3, 98)	10.48	64.56	87.57	63.42

The results obtained for the second sub-period seem to contradict the evidence that characterises the entire 20-year period and the first sub-period. During the years 2006–2014 that were characterised by extreme volatility values caused by the financial crisis originating with the Lehman Brothers bankruptcy, the VIX index and the various historical methodologies show similar forecasting ability. Therefore, it is not possible to identify the best predictive estimator. Only the exponential moving averages seem to dominate the implied volatility.

The evidence is even more surprising when we consider that the trading of option contracts on the VIX should have increased – not reduced – its information content and forecasting power.

Table 4. Regression models for the different measures of realised volatility, assuming the VIX level as an independent variable and the historical volatility computed by a GARCH (1,1) model

Dependent variables for the period 01/1995–12/2014				
	$\sigma_{Dev.std}$	σ_{Park}	$\sigma_{R\&S}$	$\sigma_{Y\&Z}$
Intercept	-0.02294**	-0.001702	0.004625	-0.0005701
	(0.0109)	(0.0086)	(0.0082)	(0.0084)

VIX_{t-1}	0.9113**	0.6530**	0.5942**	0.6558**
	(0.1083)	(0.0859)	(0.0811)	(0.0836)
$GARCH_{t-1}$	0.009463	0.02104	0.03561	0.02969
	(0.1188)	(0.0943)	(0.0890)	(0.0918)
N	227	227	227	227
Adjusted R^2	0.6117	0.5709	0.5634	0.5911
F(3, 224)	31.76	175.38	221.93	169.6
Dependent variables for the period 01/1995–02/2006				
Intercept	-0.01961	-0.01561	-0.0144	-0.01605
	(0.0188)	(0.0150)	(0.0139)	(0.0144)
VIX_{t-1}	0.8657**	0.6482**	0.5667**	0.6225**
	(0.1085)	(0.0865)	(0.0806)	(0.0832)
$GARCH_{t-1}$	0.02177	0.1552	0.2694	0.2298
	(0.2457)	(0.1959)	(0.1824)	(0.1883)
N	126	126	126	126
Adjusted R^2	0.5648	0.5692	0.5756	0.5893
F(3,123)	36.16	143.77	179.15	143.5
Dependent variables for the period 03/2006–12/2014				
Intercept	-0.007084	0.005648	0.01217	0.006873
	(0.01678)	(0.01332)	(0.01258)	(0.01301)
VIX_{t-1}	0.5568**	0.3904**	0.3096*	0.3818**
	(0.2088)	(0.1658)	(0.1565)	(0.1619)
$GARCH_{t-1}$	0.3916*	0.3078*	0.3386*	0.3303*
	(0.2299)	(0.1825)	(0.1723)	(0.1782)
N	101	101	101	101
Adjusted R^2	0.6129	0.5747	0.5623	0.5949
F(3, 98)	8.87	60.46	56.23	58.43

3.2 First research question: analysis of collinearity problems and identification of outliers

In order to deepen the understanding of the contrasting results obtained above, we deemed it necessary to analyse a potential problem of multi-collinearity that could affect the coefficients' estimates when two volatility measures are jointly used as independent variables in the same regression. To this end, we computed the Variance Inflation Factors (VIF) for the three different volatility forecasting methods (Table 5).

It is important to underscore first of all that all VIFs are lower than the critical value usually accepted, which is equal to ten. However, the sub-period 03/2006–12/2014 is characterised by VIFs that are very close to their critical value, raising doubts of a potential misinterpretation in evaluating the relative forecasting ability of different methods based on the above regressions. The significant gap between the VIFs computed before and after 2006 could potentially be related to the existence of some extreme observations that characterise the period 2006–2014, and that might have a significant influence on the relations between the different estimation methods.

Table 5. Regression models for the different measures of realised volatility, assuming the VIX level as an independent variable and the historical volatility computed by a GARCH (1,1) model

	$VIF_{VIX,SMA}$	$VIF_{VIX,EWMA}$	$VIF_{VIX,GARCH}$
01/1995–12/2014	5.195	5.713	4.897

01/1995–02/2006	2.992	3.251	2.465
03/2006–12/2014	8.239	9.384	8.646
03/2006–12/2014 *	3.481	5.287	5.316

**Sample obtained with the exclusion of observations from 09/2008 to 04/2009.*

In order to reduce the collinearity problem, we re-ran all regressions using a different sub-sample that excludes the outliers identified using the leverage influence measure and the Cook's distance applied to the original regressions. In particular, we identified seven outliers, all located at the heart of the subprime crisis between September 2008 and April 2009. Indeed, for all the variables under analysis, the new 'polished' sub-samples present considerable reductions in the VIF values that almost halve their sizes. The reduction in VIFs, excluding the volatility peak reached during the years 2008–2009, confirms our initial hypothesis that these observations have a significant impact on the relations examined.

Thus, Table 6 reports the results of the regression models that were run on the revised sub-sample in the second sub-period (2006–2014). Excluding the outliers from the data, the implied volatility dominates both GARCH and EWMA volatilities in terms of predictive power, and their contribution becomes statistically non-significant.

These results, which are more consistent with the existing literature on the topic, confirm that the volatility implied in the option prices, which directly reflect market expectations, better approximate actual market movements. That being said, our empirical evidence also highlights the inability of the VIX to correctly capture extreme market movements (as we already underlined at a descriptive level) when observing the values of RMSE in Table 1.

Table 6. Regression models for the different measurements of realised volatility for the sub-period 03/2006–12/2014, excluding the outliers

Dependent variables for the period 03/2006–12/2014				
	$\sigma_{Dev.std}$	σ_{Park}	$\sigma_{R\&S}$	$\sigma_{Y\&Z}$
Intercept	0.01214	0.02702*	0.02993**	0.02752**
	(0.01891)	(0.01425)	(0.01273)	(0.01343)
VIX_{t-1}	0.7045**	0.4147**	0.3782**	0.4357**
	(0.2068)	(0.1558)	(0.1392)	(0.1469)
$EWMA_{t-1}$	0.04052	0.1042	0.1035	0.09344
	(0.1938)	(0.1461)	(0.1305)	(0.1377)
N	95	95	95	95
Adjusted R^2	0.4139	0.3742	0.3931	0.4129
F(3,90)	16.40	95.75	135.64	99.61
Dependent variables for the period 03/2006–12/2014				
Intercept	0.01093	0.02366*	0.02666**	0.02457*
	(0.01798)	(0.01356)	(0.01213)	(0.01279)
VIX_{t-1}	0.7498**	0.4486**	0.4343**	0.4892**
	(0.2074)	(0.1565)	(0.1399)	(0.1476)
$GARCH_{t-1}$	-0.008319	0.08642	0.05656	0.04738

	(0.2440)	(0.1841)	(0.1646)	(0.1736)
N	95	95	95	95
Adjusted R^2	0.4136	0.3722	0.3897	0.4104
F(3,90)	16.38	95.35	134.72	99.07

3.3 Second research question: volatility index performance as an estimate of the market risk in a Value-at-Risk model

The last part of this paper examines the use of the VIX in a Value-at-Risk model to quantify the maximum potential loss of a hypothetical position on the S&P 500 index. A comparison of its performance against the other methods based on historical data is also conducted.

Table 7 briefly shows the key variables required to assess the performance of the various volatility parameters. The first two columns present the number of exceptions (X) and the corresponding failure rate (π), respectively.

A first reading of these variables immediately shows a significant difference between the high-volatility and low-volatility periods in terms of exceptions observed. With the sole exception of simple moving averages at a 95% confidence level, estimation methods are significantly worse in the high-volatility period than in the low-volatility one in all other cases. This evidence is also confirmed by the corresponding failure rate, which is always greater than the theoretically established one, both at the 95% and 99% confidence levels.

The picture is partly different if we only consider the second period; more precisely, while the rate obtained for the moving averages is higher at both confidence levels, those derived from the VaR based on GARCH and VIX are lower than the expected 5%, and are slightly higher considering a 1% level.

In order to obtain clearer evidence of these differences, the third column reports the statistics associated with Kupiec's unconditional coverage test (LR_{pof}), which refers to the null hypothesis of model adequacy. Its application further validates the above data for the two different periods. In the first one, only the VaR computed by simple moving averages can be considered adequate at the 95% confidence level. Nevertheless, it must be said that even just one additional exception would make the model inadequate. In all other cases, the failure rates are statistically higher, denoting the inability to accurately quantify market risk.

The results based on the second period are completely different. Only in the case of moving averages with a 99% confidence level can the null hypothesis can be rejected, indicating a strong discrepancy in the performance of the estimation methods across the different periods with the simple moving averages providing better quantification of the market risk during the most volatile phases. On the other hand, despite the worse performance during the more volatile period, the GARCH models (1.1) and the VIX present a significantly better performance in the more stable market phase.

In order to take the temporal distribution of VaR violations into account, we provide the statistics of the independence of the exceptions test (LR_{ind}) and the conditional coverage test (LR_{cc}) in Table 7. This allows us to jointly evaluate the independence and frequency of exceptions.

Regarding the independence test, it must be underlined that the test's functional form prevents the calculation of relative statistics for models without consecutive exceptions. However, in order to evaluate this property and to have a measure of the conditional coverage for each estimation method, a dummy infinitesimal empirical probability was also added to models that lacked consecutive exceptions. In this way, we distinguish it from zero even if this value is basically negligible.

Despite the possibility of having a value for all the VaRs, this procedure provides misleading results in the specific case of the simple and exponential moving averages in the high-volatility period only considering the violations following a non-violation day due to the comparison between the infinitesimal given rate and the empirical observed one,. In all cases except for this singular one, and for both periods, the VaR models adequately consider the changes in market conditions and in most cases present an LR_{ind} value that is considerably lower than the critical one.

Indeed, combining the results obtained by this test and Kupiec's test, it is evident that the most significant component in determining the inadequacy of models is the high empirical failure rate, especially in the high-volatility period characterised by more extreme market movements. This allows us to observe that even in the more volatile phases, all the methods used to quantify the maximum potential loss provide a measure that, despite the high number of exceptions, seems to adequately handle the volatility clustering that could be particularly pronounced in the bear market phases.

In order to have a complete representation of predictive performance, the last few columns of Table 7 report the key values of the Lopez loss function test. This test has confirmed that the worst performances were in the period 2008–2009, with substantial differences in losses resulting from the VIX and GARCH estimates compared to the moving average estimator ones that exhibit the best performance. These results are relatively surprising. In fact, we expected both GARCH and VIX to be able to better forecast market volatility in turbulent market conditions, since the GARCH model takes volatility clustering into account, while the Volatility Index, known as the 'investor fear gauge', is based on the market expectations and starts to rise during times of financial stress.

The evidence described however reverses completely in the low-volatility period where the size of the losses of these two methods is significantly lower than those obtained through the moving averages estimates.

Very similar evidence is achieved when the number of violations on the variable described above is isolated by the use of the average size of losses. In this sense, the average size of violations made by VIX in the first period is noteworthy, as it is more than twice the GARCH one. In a broad sense, the smaller average size of losses calculated by the GARCH model, when compared with the VIX's average size of losses seems to indicate that even if the VIX provides a better measure in terms of failure rate, it leads to significant losses when exceptions occur.

In conclusion, analysis of these different methods for predicting the market risk shows that the simple moving averages perform better during the most critical period. This evidence is quite surprising and probably due to the short time frame in which they are calculated, which allows them to react faster to new market shocks and contain the fraction of losses exceeded by the VaR.

The use of the VIX and GARCH volatiles instead, seems to be inadequate in such a context for both the number of exceptions and the amount of losses. Conversely, their performance is significantly better when the market faces more normal conditions, where the lower volatility allows them to reduce the corresponding failure rate and the average losses that occur.

Table 7. Tests to evaluate the performance of the different volatility estimators in the VaR calculation

High-volatility period 2008-2009								
		X	π	LR_{pof}	LR_{ind}	LR_{cc}	C_M	$\overline{C_M}$
Confidence level 95%	VAR_{SMA}	35	6.93%	3.60	5.22**	8.82**	-35.22%	-1.01%
	VAR_{EWMA}	38	7.40%	5.96**	6.20**	12.16**	-36.25%	-0.95%
	VAR_{GARCH}	44	8.71%	12.19**	3.40	15.59**	-48.27%	-1.10%
	VAR_{VIX}	46	9.11%	14.68**	0.45	15.13**	-51.10%	-1.11%
Confidence level 99%	VAR_{SMA}	14	2.77%	10.85***	0.80	11.65***	-8.75%	-0.62%
	VAR_{EWMA}	14	2.77%	10.85***	0.80	11.65***	-7.93%	-0.57%
	VAR_{GARCH}	21	4.16%	28.53***	1.82	30.35***	-13.97%	-0.67%
	VAR_{VIX}	12	2.38%	7.00***	0.58	7.58***	-16.80%	-1.40%
Low-volatility period 2013–2015								
		X	π	LR_{pof}	LR_{ind}	LR_{cc}	C_M	$\overline{C_M}$
Confidence level 95%	VAR_{SMA}	35	7.00%	3.81	0.10	3.91	-16.25%	-0.46%
	VAR_{EWMA}	27	5.41%	0.17	3.09	3.26	-14.79%	-0.55%
	VAR_{GARCH}	22	4.40%	0.38	2.03	2.41	-9.21%	-0.42%
	VAR_{VIX}	20	4.00%	1.11	1.67	2.78	-9.78%	-0.49%
Confidence level 99%	VAR_{SMA}	13	2.60%	9.01***	0.70	9.71***	-6.96%	-0.54%
	VAR_{EWMA}	13	2.60%	9.01***	0.70	9.71***	-5.66%	-0.44%
	VAR_{GARCH}	7	1.40%	0.73	0.19	0.92	-1.93%	-0.28%
	VAR_{VIX}	8	1.60%	1.55	0.26	1.81	-2.93%	-0.37%

*Where X is the number of exceptions, π is the empirical exception rate, LR_{pof}, LR_{ind} and LR_{cc} are, respectively, the key statistics for the Kupiec's test, independence test and the conditional coverage test. ** and *** indicate the rejection of the null hypothesis with confidence level of 95 and 99%, and thus, the inadequacy of the VAR model. C_M is the value taken by the loss function and $\overline{C_M}$ is the average size of this function.*

4 Conclusion

The first aim of this study was to investigate whether the Volatility Index (VIX) is able to predict future realised volatility and to examine what the corresponding information content is. In line with the mainstream literature on this topic, our results point out that the VIX is a biased estimator of realised volatility, although its ability to explain a considerable portion of realised performance allows it to dominate the other methods based on historical data. Despite the possibility of taking a direct stand in terms of expected volatility by introducing options contracts written on the VIX, the information content of this index has not significantly changed. By directly analysing the predictive power of the VIX against other methods based on historical data, the superiority of the VIX is confirmed in the entire period and before 2006. Contrastingly in the second period after 2006 that is characterised by higher volatility due to the financial crisis, the results are not clear-cut. These differences between the two sub-periods prompted us to deepen our analysis. In particular, we found collinearity issues that affect the results in the period 2006–2014, and are basically caused by the presence of some abnormal observations during the most volatile market phase that started in September 2008. Indeed, leaving out these outliers, the empirical evidence is consistent with previous studies, confirming the better predictive power of the VIX.

The second aim of this paper concerned the ability of the VIX to quantify the maximum potential loss. As discussed in the previous sections above, the backtests performed for the VaR model show a poor performance of VIX during the most turbulent market phase, thus causing the inadequacy of this model both for the failure rate and the size of losses. The preferable performance of simple moving averages can be noted instead.

On the contrary, in periods characterised by remarkably lower VIX levels, its performance significantly improves, considerably limiting the number of violations and the magnitude of losses.

Despite the very similar performance of the VIX and GARCH models, the latter method seems preferable, as proved by a failure rate that is closer to the theoretical one, and also by the minimisation of VaR losses. However, in the period 2008–2009, GARCH shows a higher exception rate than the VIX with a 99% confidence level, but is associated with a much lower value of the average entity of losses.

References

Becker, R., Clements, A.E. and White, S.I. (2007), 'Does implied volatility provide any information beyond that captured in model-based volatility forecasts?', *Journal of Banking and Finance,* 31, pp. 2535-2549.

Bollersev, T. (1986), 'Generalized autoregressive conditional heteroschedasticity'*, Journal of Econometrics,* 31, 307-327.

Brownless, C., Engle, R. and Kelly, B. (2011), 'A practical guide to forecasting through calm and storm', *The Journal of Risk,* 14(2), 2011, pp. 3-22.

Canina, L. and Figlewski, S. (1993), 'The informational content of implied volatility', *The Review of Financial Studies*, 6(3), pp. 659-681.

Charoenwong C., Jenwittayaroje N. and Sin Low B. (2009), 'Who knows more about future currency volatility?', *The Journal of Futures Markets,* 29(3), pp. 270-295.

Chicago Board Options Exchange (2014), '*The CBOE Volatility Index – VIX. The powerful and flexible trading and risk management tool from the Chicago Board Options Exchange'*, Chicago: CBOE.

Chong, J. (2004), ‘Value at Risk from econometric models and implied from currency options’, *Journal of Forecasting,* 23, pp. 603-620.

Christensen, B.J. and Prabhala, N.R. (1998), ‘The relation between implied and realized volatility’, *Journal of Empirical Finance*, 50, pp. 125-150.

Christoffersen, P., Hahn, J. and Inoue, A. (2001), ‘Testing and comparing Value-at-Risk measures’, *Journal of Empirical Finance*, pp. 325-342.

Christoffersen, P. F. (1998), ‘Evaluating interval forecasts’, *International Economic Review,* pp. 841-862.

Corrado, C. J., & Miller Jr and T. W. (2005), ‘The forecast quality of CBOE implied volatility indexes’, *Journal of Futures Markets,* 25(4), pp. 339-373.

Day, T.E. and Lewis, C.M. (1992), ‘Stock market volatility and the information content of stock index options’, *Journal of Econometrics,* 52, pp. 267-287.

Ederington, L. and Guan, W. (2002), ‘Is implied volatility an informationally efficient and effective predictor of future volatility?’ *Journal of Risk,* 4, pp. 29-46.

Fleming, J. (1998), ‘The quality of market volatility forecasts implied by S&P 100 index option prices’, *Journal of Empirical Finance,* 5, pp. 317-345.

Giot, P. (2005), ‘Implied volatility indexes and daily Value at Risk models’, *The Journal of Derivatives,* 12(4), pp. 54-64.

Jorion, P. (1995), ‘Predicting volatility in the foreign exchange market’, *The Journal of Finance,* 50, pp. 507-528.

Kim, J. S. and Ryu, D. (2015), ‘Are the KOSPI 200 implied volatilities useful in value-at-risk models?’, *Emerging Markets Review,* 22, pp. 43-64.

Kourtis, A., Markellos, R.N. and Symeonidis, L. (2016), ‘An international comparison of implied, realised and GARCH volatility forecasts’, *The Journal of Futures Markets,* (0), pp. 1-30.

Kupiec, P. (1995), ‘Techniques for verifying the accuracy of risk measurement models’*, Journal of Derivatives,* 2, 173-184.

Lopez, J. A. (1999), ‘Methods for evaluating value-at-risk estimates’, *Economic Review,* 2, pp. 3-17.

Parkinson, M. (1980), ‘The extreme value method for estimating the variance of the rate of return’, *Journal of Business*, pp. 61-65.

Poon, S.H. and Granger, C. (2005), ‘Practical issues in forecasting volatility’, *Financial Analyst Journal,* 61(1), pp. 45-71.

Shu, J. and Zhang, J. E. (2003), 'The relationship between implied and realized volatility of S&P 500 index', *Wilmott Magazine,* 4, pp. 83-91.

Szakmary, A., Ors, E., Kim, J. K. and Davidson, W. N. (2003), 'The predictive power of implied volatility: Evidence from 35 futures markets', *Journal of Banking & Finance,* 27(11), pp. 2151-2175.