International Journal of Forecasting 33 (2017) 848–863

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Volatility measures and Value-at-Risk

Dennis Bams [a,b], Gildas Blanchard [a], Thorsten Lehnert [c,*]

[a] Limburg Institute of Financial Economics (LIFE), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[b] Open University Heerlen, P.O. Box 2960, 6401 DL Heerlen, The Netherlands
[c] Luxembourg School of Finance, University of Luxembourg, 4, rue Albert Borschette, L-1246 Luxembourg, Luxembourg

ARTICLE INFO

Keywords:
Value-at-Risk
Option implied volatility
Volatility risk premium
Time-series
GARCH models

ABSTRACT

We evaluate and compare the abilities of the implied volatility and historical volatility models to provide accurate Value-at-Risk forecasts. Our empirical tests on the S&P 500, Dow Jones Industrial Average and Nasdaq 100 indices over long time series of more than 20 years of daily data indicate that an implied volatility based Value-at-Risk cannot beat, and tends to be outperformed by, a simple GJR-GARCH based Value-at-Risk. This finding is robust to the use of the likelihood ratio, the dynamic quantile test or a statistical loss function for evaluating the Value-at-Risk performance.

The poor performance of the option based Value-at-Risk is due to the volatility risk premium embedded in implied volatilities. We apply both non-parametric and parametric adjustments to correct for the negative price of the volatility risk. However, although this adjustment is effective in reducing the bias, it still does not allow the implied volatility to outperform the historical volatility models.

These results are in contrast to the volatility forecasting literature, which favors implied volatilities over the historical volatility model. We show that forecasting the volatility and forecasting a quantile of the return distribution are two different objectives. While the implied volatility is useful for the earlier objective function, it is not for the latter, due to the non-linear and regime changing dynamics of the volatility risk premium.

1. Introduction

Numerous empirical works have demonstrated the superiority of option-implied volatility (IV) over historical volatility models for predicting the future return volatility. We contribute to this literature by evaluating and comparing the merits of IV and historical volatility models for Value-at-Risk (VaR) forecasting. Volatility forecasting and VaR forecasting are two different objectives. Hence, this paper extends the comparison between IV and time series information into another field. Furthermore, we show that the best volatility forecast is not the best VaR forecast. The results of our multiple and complementary back-testing procedures show that the IV-based VaR cannot outperform a standard historical volatility VaR model.

The non-linear and regime changing dynamics of the volatility risk premium embedded in option prices explains the disappointing performance of IV for VaR forecasting. The risk neutral expected volatility implied from options differs from the physical volatility. Both IV-based volatility forecasts and VaR forecasts need to be corrected to address this difference. However, in the case of volatility prediction, a simple correction is sufficient to transform IV into the best forecast. On the other hand, the simple corrections applied in this paper do not allow the IV-based VaR to outperform the historical volatility model based VaR. The quantile prediction exercise concentrates on forecasting the occurrence of tail events. The complex dependence structure between the volatility risk premium and extreme

* Corresponding author.
E-mail address: thorsten.lehnert@uni.lu (T. Lehnert).

http://dx.doi.org/10.1016/j.ijforecast.2017.04.004
0169-2070/

returns affects the quantile forecasting power of IV adjusted measures.

Despite earlier studies indicating the poor information content of IV (Canina & Figlewski, 1993), the great majority of volatility prediction literature has concluded that IV is either superior to historical volatility model forecasts (Blair et al., 2010; Christensen & Prabhala, 1998; Fleming, 1998), or at least complements it (Beckers, 1981; Day & Lewis, 1992). This conclusion has been reached based on a variety of markets, including: equity indices (Corrado & Miller, 2005; Yu et al., 2010), individual equities (Taylor et al., 2010), currencies (Charoenwong et al., 2009) and commodities (Szakmary et al., 2003). Only intraday time series volatility information appears to compete with IV information (Pong et al., 2004; Taylor & Xu, 1997).

Many of the studies in this field justify their investigations in the name of risk management (Frijns et al., 2010; Martens & Zein, 2004). However, the earlier literature concentrated on volatility forecasting rather than on its risk management applications. This paper evaluates whether the IV's superior volatility forecasting power translates into a superior VaR prediction relative to historical volatility models. More recently, a stream of studies has investigated the benefits of IV in the context of VaR forecasting (for a comprehensive survey, see Nieto & Ruiz, 2016). Despite this growing body of evidence, no consensus has been reached. Giot (2005) and Jeon and Taylor (2013) show that, for equity indices, a combination of IV and time series information results in superior VaR predictions. For currency markets, the findings of Chong (2004) and Christoffersen and Mazzotta (2005) suggest that IV models provide poor forecasts of the tail of the returns distribution.

We identify the volatility risk premium as a key component, which could explain the discrepancies in the conclusions reached. The relative size of the volatility risk premium across markets, such as equity and currency, is a distinctive factor. Furthermore, the various methodological choices provide implicit adjustments for the volatility risk premium. For instance, incorporating IV into the conditional volatility (quantile) equation, or aggregating IV with other forecasts, provides a source of non-transparent adjustment. Therefore, the IV-based VaR forecast performance is very sensitive to the methodological settings adopted in various investigations.

The empirical design adopted here differs from those of more recent studies that have been dedicated to the use of IV in a VaR forecasting exercise. First, we can account for and isolate the volatility risk premium affecting the IV-based forecasts formally. This approach provides transparent information and quantifies the impact of the volatility risk premium. It also allows us to distinguish clearly the relative benefits of standalone IV-based forecasts. Second, we perform a formal evaluation of the relative performances of IV models and historical volatility models. Previous studies on the topic have relied on traditional back-testing tests, which evaluate the individual forecasting performances but cannot establish the statistical superiority between two competing models.

IV-based forecasts, although efficient, have been documented to be biased (Lamoureux & Lastrapes, 1993; Szakmary et al., 2003; Yu et al., 2010). Chernov (2007) formally identifies the origin of this bias, namely the volatility risk premium (VRP). Since the price of volatility risk is negative, the risk neutral expectation of the volatility implied from options is higher than the physical expectation of that is volatility relevant for the VaR calculation. This feature has a negative effect on the predictive power of IV-based models (Tsiaras, 2009). Prokopczuk and Wese Simen (2014) were the first to formally acknowledge and account for the variance risk premium in an assessment of the forecasting performance of IV. They demonstrate empirically that a simple adjustment or correction of implied volatilities for the variance risk premium improves the forecasting performance relative to standard IV models both in and out-of-sample.

We account for the volatility risk premium in the same spirit as Prokopczuk and Wese Simen (2014), but apply the procedure in the VaR prediction context instead. Moreover, we focus on three major equity indices with publically available implied volatility indices, rather than the commodity market, which requires a proprietary option dataset. The behaviours of the IV and the volatility risk premium differ across markets and assets (Martin et al., 2009), meaning that it is relevant to evaluate the performances of adjusted IV forecasts on more mainstream markets, such as major equity indices.

We assess the information content, for the purpose of VaR measurement, of historical volatility models, IV and IV-adjusted for the volatility risk premium over an extended time period. We adjust for the volatility risk premium both non-parametrically and parametrically, and compare the results obtained. We measure the VaR for three major stock indices, S&P 500, DJIA and Nasdaq 100, using three implied volatility indices SPX (1990–2013), VXD (1997–2013) and VXN (2001–2013). We study the 95% and 99% confidence level VaR, as well as the one day out-of-sample and one month out-of-sample VaR. The one month VaR is included in this study in order to match the maturity of the implied volatility indices, even though it is not used in practice.

We evaluate the VaR models both individually and relative to each other. Individual performances are assessed using traditional backtesting tests. We use both the LR test of Kupiec (1995), which measures whether the number of exceptions is consistent with the confidence level, and the dynamic quantile test of Engle and Manganelli (2004). Relative forecasting performances are established using a VaR specific loss function. The loss function's statistical difference is obtained through a bootstrapping approach, as per Chen and Gerlach (2013), and a formal conditional predictability test proposed by Giacomini and White (2006).

The results from the volatility forecasting literature cannot be transposed to a quantile forecast application. Although IV information generally outperforms time series information for volatility prediction, our results suggest that such is not the case for VaR forecasting: IV and IV-adjusted VaR do not outperform GJR-GARCH VaR. In fact, under certain circumstances, the latter historical volatility model is even found to outperform all of the other models. The GJR-GARCH passes the LR and dynamic quantile tests, at the 5% significance level, more successfully than the other VaR models, including the IV-based

VaR. A model-to-model statistical loss function comparison shows that the GJR-GARCH always belongs to the group of best models. The failure of IV to provide a better VaR than the historical volatility model is robust to changes in the VaR confidence level and the market under investigation. The maturity mismatch between the VaR horizon and the option maturity cannot explain these results.

We show that the poor performance of IV-based VaR forecasts is due to the volatility risk premium, embedded in option prices. Because equity indices' IV is rationally higher than the realized volatility, the IV-based VaR severely overstates the actual VaR. This translates into an insufficient number of observed violations for the three indices. An inaccurate, excessively conservative VaR implies an excess capital charge for financial institutions. When options IV are used to forecast VaR, there is a need to account for the volatility risk premium.

Relatively simple IV adjustments correct and improve the volatility forecasting power of option implied information significantly (DeMiguel et al., 2013; Prokopczuk & Wese Simen, 2014). Although the parametric and non-parametric volatility risk premium adjustments that we propose reduce the downward bias of the standard VaR IV, they do not allow IV to outperform the GJR-GARCH model. While standard volatility risk premium adjustments are successful for volatility forecasting, such does not hold for quantile prediction. These adjustments assume a parsimonious linear relationship between the volatility risk premium, or relative volatility risk premium, and the volatility level. This parsimonious specification is enough to capture the future average volatility properly. However, a richer dynamic is necessary to predict the tail of the return distribution. Our results and other evidence found in the literature suggest that the relationships between the volatility risk premium, volatility, returns and innovations are highly non-linear around extreme events. Therefore, the linear assumptions are the most harmful for predicting extreme quantiles of the return distribution.

We conclude that the use of option-implied information for risk management purposes is not as efficient as the volatility forecasting literature has commonly supposed. The volatility risk premium included in option-implied information and the adjustment technique used do not provide sufficiently good VaR forecasts to beat simple historical volatility model VaR forecasts.

2. Data

We proxy the model-free IVs of the S&P 500, DJIA and Nasdaq 100 using their respective Chicago Board of Exchange (CBOE) volatility indices. They represent the investor's risk-neutral expectation of the volatility over the next 30 days for these three market indices. We rely on the new volatility indices that are computed from the whole range of options available, rather than only the at-the-money options, as was previously the case. This approach accounts for the implied volatility smile and skew featured by the option market. Corrado and Miller (2005) demonstrate empirically that the VXO, VIX and VXN are informationally efficient and that the notable error-in-variable problem for these indices is no longer a real concern after 1995. These volatility indices are computed without relying on any model, such as that of Black and Scholes (1973), and are therefore truly model-free measures. Martin et al. (2009) show that the VIX replicates the model-free measure of Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) very precisely.

We collect daily return series on the underlying equity indices from DataStream. The data cover different periods for the three indices, depending on the availability of the volatility indices, which varies across indices. The sample covers the S&P 500 from 1990, the DJIA from 1997 and the Nasdaq 100 from 2001, all until 2013. Fig. 1 shows some obvious common trends across the three indices' price levels and IVs. However, the three markets display different levels of volatility, with Nasdaq 100 being the most volatile index. Table 1 compares the indices' volatilities more formally. Both the 22-day realized volatilities and the IVs are presented. Considering the time period covered for all three markets (after 2002), the realized annualized volatility of the Nasdaq 100 is 4.7% (5.9%) higher than the S&P 500 (DJIA) volatility. The DJIA is more volatile than the S&P 500 during the earliest part of our sample, but the situation reverses after 2002.

The next section evaluates the benefits of using IV, a risk neutral measure, to form VaR forecasts for actual physical returns. Table 1 provides statistics regarding the spread between the risk neutral volatility and physical volatility. As has been documented widely, the volatility risk premium is negative and the IV surpasses the subsequent realized volatility (Carr & Wu, 2009; Mixon, 2009). The average size of the premium is economically significant, ranging from 3.26% to 5.51%. These premiums are also statistically significant. All of the mean volatility risk premiums reported in Table 1, except for the Nasdaq 100 before 2002, are significant at the 1% level using Newey–West standard errors.

In the literature, the negative price of the variance risk is traditionally computed as the difference between the variance under the real-world ($\mathbb{P}$) and the risk-neutral ($\mathbb{Q}$) expectations. However, the remainder of this paper approaches the premium slightly differently. First, the premium is expressed in term of standard deviations rather than variances. This allows the values to be compared with the observed returns directly. Second, the premium is computed as $\mathbb{Q}$–$\mathbb{P}$, in order to obtain mainly positive figures. This is in line with the definition that is used in more recent studies, e.g., Bali and Zhou (2016). The underlying dynamics are identical, despite this difference in notation.

The volatility risk premium is known to be a positive function of the volatility level (Carr & Wu, 2009; Martin et al., 2009). However, we observe that although Nasdaq 100 is by far the most volatile market, it does not have a significantly higher volatility risk premium. Also, the Nasdaq 100 volatility risk premium is more volatile than those of the other indices. Using a smaller sample, Corrado and Miller (2005) also showed that the Nasdaq 100 spread between IV and the realized volatility was almost non-existent, and meaningfully lower than for other indices. They interpret a good predictive power of the VXN over the

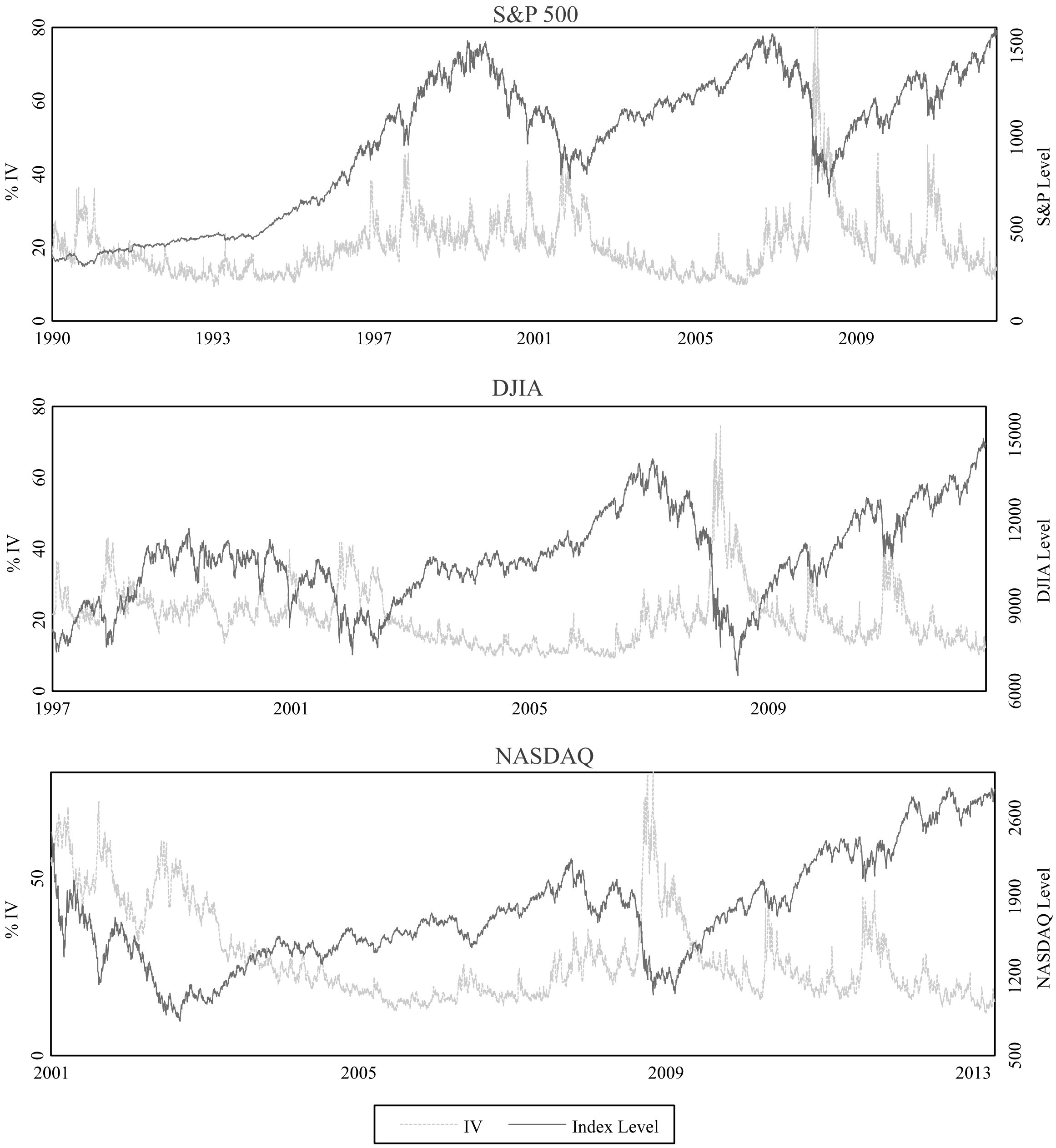

Fig. 1. Indices and implied volatility levels. Note: The figures show the time series of the index in dark solid grey and the implied volatility in light dashed grey, for the three indices under investigation: S&P 500, DJIA and Nasdaq 100. The time series are not all the same lengths because we are limited by the availability of the volatility indices. The time series only starts when the implied volatility is available. This explains why S&P 500 has a longer time series than Nasdaq 100: the VIX is available from 1990 onwards, while the VXN is only available after 2001.

subsequent realized volatility as a sign of the Nasdaq 100 option market's efficiency, but it can also be explained by the relatively smaller volatility risk premium making IV a less upward biased estimator for that specific market.

Aside from the economic importance of the price of the volatility risk, Table 1 highlights the fact that the raw IV is not an appropriate input for a VaR forecast, because of the volatility risk premium. Thus, an adjustment is required, as was suggested by DeMiguel et al. (2013), Jorion (1995), or Prokopczuk and Wese Simen (2014).

3. Methodology

The VaR at the $(1-\alpha)$ % confidence level measures the loss of an investment that should not be exceeded (worst loss) in more than α% of cases. VaR forecasting boils down to predicting the α% quantile of the return distribution. The one-day VaR is used for our empirical application, though a one-month horizon VaR is also tested for robustness purposes.

The Basel Committee on Banking Supervision made capital requirements for financial institutions dependent

Table 1
Annualized implied volatility, realized volatility and the volatility risk premium.

	S&P				DJIA				NASDAQ			
	Mean	STDev	5%	95%	Mean	STDev	5%	95%	Mean	STDev	5%	95%
Total sample												
IV	20.4	8.1	11.5	34.8	21.0	8.3	11.0	36.5	28.2	13.0	15.1	56.4
RV	16.0	9.4	7.1	32.1	17.1	9.7	7.5	36.0	24.4	14.2	11.0	53.6
VRP	4.4	6.0	−4.0	11.7	3.9	6.6	−6.4	11.7	3.8	7.6	−7.8	13.3
Before 2002												
IV	19.5	6.1	11.6	30.5	24.2	4.6	18.4	33.1	54.6	7.2	45.4	66.2
RV	14.3	6.4	6.5	27.3	18.7	6.6	11.6	33.7	49.4	14.0	32.1	79.8
VRP	5.2	4.7	−27.0	12.1	5.5	6.6	−8.7	13.6	5.2	11.2	−16.7	20.9
After 2002												
IV	21.3	9.7	11.4	40.7	19.8	9.0	10.8	37.8	26.0	10.9	15.0	48.7
RV	17.7	11.6	7.8	39.7	16.5	10.6	7.3	37.0	22.4	12.2	10.9	47.5
VRP	3.6	7.1	−6.3	11.5	3.3	6.5	−5.6	10.8	3.7	7.3	−6.1	12.1

Note: The table represents the implied volatilities from the VIX, VXD and VXN, the realized volatilities from the S&P 500, the DJIA, and the Nasdaq 100. The variables are collected or computed on a daily basis. The sample covers 23, 16 and 12 years of data for the S&P 500, DJIA and Nasdaq 100, respectively. The realized volatilities are computed as the sum of the 22 subsequent squared log-returns. The volatility risk premium (VRP) is computed as implied volatility (IV) minus realized volatility (RV). All figures are annualized and reported as percentages.

on the VaR. Thus, the accuracy of the VaR is critical in order to ensure that they have sufficient, but not excessive, capital. For that reason, alternative VaR models have been evaluated both in practice and in academia (Berkowitz & O'Brien, 2002; Kuester et al., 2006). This paper contributes to this body of literature by investigating the benefits of using option-implied information to measure the VaR.

We assume that returns at time t, r_t, are defined as $r_t = \sigma_t \times \varepsilon_t$, where σ_t is the volatility at time t and ε_t is the innovation at time t. We obtain different forecasts of σ_t using IV and historical volatility information alternatively, and compare the resulting VaR performances. We avoid making any assumptions about the distribution of the innovations and ensure similar treatments of the innovations across all models by applying a data driven approach, as described below.

3.1. Time series based forecasts

The first forecast that we use for σ_t is the usual past realized volatility over the last 22 trading days, as defined by Jorion (1995):

$$\sigma_t^{RV} = \sqrt{\frac{1}{22}\sum_{i=t-22}^{t-1} r_i^2}, \tag{1}$$

where r_i^2 is the squared daily return at time i assuming that the mean daily return is equal to zero, and σ_t^{RV} represents the forecast of σ_t.

The second time series based approach that we rely on is the forecast of the GJR-GARCH model developed by Glosten et al. (1993). This volatility model is known to provide a good description of the volatility dynamics (Engle & Ng, 1993; Rosenberg & Engle, 2002). The following equations describe the GJR volatility model specifications:

$$r_t = \sigma_t \times \epsilon_t \tag{2}$$

$$\sigma_t^2 = \omega + \alpha\, r_{t-1}^2 + \gamma\, r_{t-1}^2\, I_{[r_{t-1}<0]} + \beta\, \sigma_{t-1}^2, \tag{3}$$

where r_t is the daily return at time t and ϵ_t is a standardized normally distributed innovation at time t. As in a standard GARCH(1,1) model, the conditional variance depends on previous returns and the previous conditional variance. Moreover, the indicator variable I takes a value of 1 when the previous return is negative and 0 otherwise. This additional term accounts for the leverage effect[1] that characterizes equity returns (Hansen & Lunde, 2005). The Matlab-based toolbox MFE was used to perform some of the time series estimations and predictions.[2]

We formed a GJR volatility forecast σ_t^{GJR} by estimating the model parameters over a rolling window of 500 observations up to $t-1$, and computed the out-of-sample σ_t^{GJR} based on the following equation for a one-day volatility forecast:

$$\widehat{\sigma}_t^{GJR} = \sqrt{\widehat{\omega} + \widehat{\alpha}\, r_{t-1}^2 + \widehat{\gamma}\, r_{t-1}^2\, I_{[r_{t-1}<0]} + \widehat{\beta}\, \sigma_{t-1}^2}. \tag{4}$$

Eq. (4) provides a one-day-ahead forecast, but forecasts over n days can also be obtained by summing the n individual daily conditional variances. We compute GJR-GARCH monthly volatility forecasts by summing 22 daily conditional variances according to the following equation[3]:

$$\widehat{\sigma}_{t+n}^{GJR} = \sqrt{\widehat{\sigma}^2 + \left(\widehat{\alpha} + 0.5\widehat{\gamma} + \widehat{\beta}\right)^{n-1}\left(\widehat{\sigma}_t^2 - \widehat{\sigma}^2\right)}, \tag{5}$$

where $\widehat{\sigma}^2$ is the unconditional variance of returns, computed as $\widehat{\omega}/(1-(\widehat{\alpha}+0.5\widehat{\gamma}+\widehat{\beta}))$, $\widehat{\sigma}_{t+1}^2$ is the first one-day-ahead conditional variance (provided by Eq. (4)) squared, and n is equal to 22. The parameter $\widehat{\gamma}$ is multiplied by 0.5, relying on the assumption that returns follow a symmetric distribution. Accordingly, returns have the same probabilities of being gains or losses, and the expectation of the indicator variable $I_{[r_{t-1}<0]}$ is 0.5. Under stationary conditions ($1 > \widehat{\alpha} + 0.5\widehat{\gamma} + \widehat{\beta}$), daily conditional variances will

[1] Our initial tests justify the use of a GJR-GARCH rather than a simpler volatility specification without leverage.

[2] https://www.kevinsheppard.com/MFE_Toolbox.

[3] For a formal derivation of the forecast equation, together with detailed information regarding unconditional variance, conditional variance and persistence parameters for the GJR-GARCH model, see Taylor (2005).

converge in time geometrically toward the unconditional variance of the process at a rate that is defined by the values taken by the parameters $\widehat{\alpha}, \widehat{\gamma}$ and $\widehat{\beta}$. As a consequence, the GJR-GARCH volatility is smoothed over long-horizon forecasts.

3.2. Implied volatility based forecasts

The first IV forecast simply uses the observed IV from the implied volatility indices at the end of day $t-1$. Because the volatility indices are computed for 30-day maturity options that are commonly assumed to be 22 traded days, we adjust for the maturity as follows:

$$\widehat{\sigma}_t^{IV} = \sqrt{\frac{1}{22}} \times IV_{t-1}, \tag{6}$$

where the IV_{t-1} are the VIX, VXD and VXD values at the end of day $t-1$, expressed as monthly volatilities.

This IV forecast does not account for the discrepancy between $\mathbb{P}$ and $\mathbb{Q}$ volatility expectations, but assumes that the two measures are equivalent. We adjust IV to correct for this difference by dividing IV_t by an estimate of the ratio of the implied volatility to the realized volatility (VRP_t). This ratio is estimated both parametrically and non-parametrically. The non-parametric adjustment is based on the historical median value of the ratio between the IV and the subsequent realized volatility, computed as:

$$\widehat{VRP}_t(NPA) = IV_t \Bigg/ \sqrt{\sum_{i=t+1}^{t+22} r_i^2}. \tag{7}$$

Prokopczuk and Wese Simen (2014) propose the use of the ratio instead of the difference in order to account for the well-known fact that the variance risk premium is a function of the volatility level (Carr & Wu, 2009; Martin et al., 2009). We compute the median of VRP_t over the last year (252 days). This window is a good compromise for capturing the time variation in the volatility risk premium and estimating it sufficiently well. Since we want to construct purely out-of-sample VaR measures, the last VRP ratio that is used to compute the median at time t is VRP_{t-22}. This non-parametrically adjusted IV forecast is denoted $\widehat{\sigma}_t^{IV}(NPA)$.

This non-parametric adjustment is similar to the methodology provided by DeMiguel et al. (2013) and Prokopczuk and Wese Simen (2014). These two studies use the average relative ratio of IV over the realized volatility during the last year. We differ in that we rely on the median to reduce the impact of outliers.[4]

The alternative parametric adjustment estimates the implied volatility over a realized volatility ratio from an AR(1) model:

$$VRP_t = \varphi_0 + \varphi_1 VRP_{t-22} + e_t, \tag{8}$$

where φ_1 represents the AR(1) coefficient. This model is estimated on the last 252 available observations (one year) and provides an out-of-sample $\widehat{VRP}_t$ prediction. This forecast is used to construct the parametrically adjusted IV forecast $\widehat{\sigma}_t^{IV}$ (NPA). The AR(1) coefficient is estimated to be significant for all three indices at the 5% level.[5] It could be argued that an ARMA(1,1) such as that used by Prokopczuk and Wese Simen (2014) describes the data better. Indeed, the Schwartz criterion favours the ARMA(1,1) model.[6] However, the ARMA(1,1) model, which tends to fit noise, provides some extreme predictions that cause a severe deterioration in the loss function used for VaR model evaluation.[7] However, in order to alleviate any concerns, the robustness section replaces the AR(1) model with an ARMA(1,1) model, and we find that the results and conclusions remain unchanged and unaffected by this choice of specification.

The five volatility forecasts that we use for estimating the VaR differ from each other with regard to the nature of the information used (time series or option implied information), the parametric assumption made (parametric or non-parametric) and the treatment of the volatility risk premium applied (adjusted or non-adjusted). The remainder of the paper empirically evaluates and benchmarks the performances of those approaches for providing accurate VaR measures.

So far, we have not provided any information regarding our treatment of innovations, since we only consider volatility forecasts. The performance of a VaR is affected strongly by the choice of an appropriate distribution for the innovations (Giot & Laurent, 2003; Mittnik & Paolella, 2000). It is necessary to include some form of non-normality and negative skewness in the innovation distribution in order to obtain a reliable VaR.

We avoid making any arbitrary distributional choice by relying on the empirical innovation distribution. Barone-Adesi et al. (1998) proposed the filtered historical simulation technique to account for the non-normality of innovations. In the VaR context, Kuester et al. (2006) show that filtered historical simulation is an effective way to account for the non-normality of innovations. In a similar spirit, we compute the conditional innovations by dividing the realized returns by the forecasted volatility of each model. The quantiles (α) of the previous 500 empirical innovations are used to form the current VaR. This approach, which improves the performance of the VaR drastically relative to a naive normality assumption, does not make any assumptions about the innovation distribution,[8] and is applied consistently across all models.

Filtered historical simulations can be applied for all VaR candidates, including the non-parametric models. Parametric forms of innovations generally require historical data and parametric estimations, and therefore can only be used for the VaR^{GJR} model, not applying to the other

[4] We tried using multiple different horizons and time windows to compute the median VRP ratio, but the results are insensitive to these choices.

[5] The coefficient is estimated to be significant for more than 58% of the estimations for all indices.

[6] The ARMA (1,1) minimizes the Schwartz criterion in more than 66% of the estimations for all indices.

[7] An investigation beyond one lag is not appropriate in this exercise, as each lag reduces the sample by 22 observations.

[8] The only implied assumption is that the innovation distribution is not time-varying.

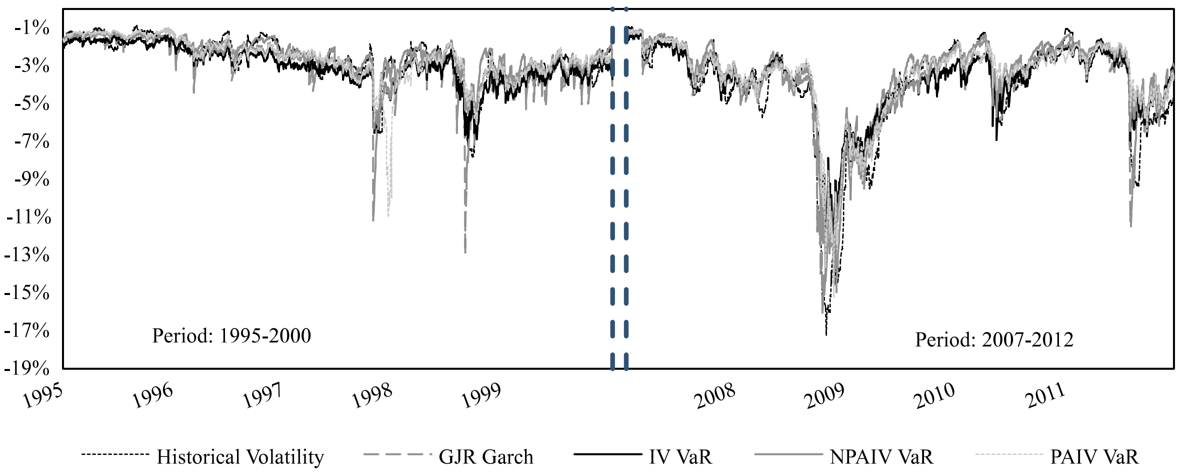

Fig. 2. VaR forecasts for the S&P 500. Note: The figure shows the VaR time series computed using different models. For graphical reasons, the graph focuses on two representative time periods (period 1: 1995–2000; period 2: 2007–2012), but VaR forecasts are available for the entire sample.

candidates. We confirm that filtered historical simulations are treating innovations efficiently by replicating the results for the VaR^{GJR} model with skewed-t distributed innovations (Hansen, 1994).

4. Results

We apply the methodology described in the previous section to obtain five out-of-sample VaR measures: VaR_t^{HV}, VaR_t^{GJR}, VaR_t^{IV}, $VaR_t^{IV}(NPA)$ and $VaR_t^{IV}(PA)$. This section analyses the properties and performances of those measures. Fig. 2 graphs the VaR levels for these different measures over two representative time periods.

VaR measures vary with respect to their smoothness and levels. Because of the extra weights that are applied to the most recent observations, the VaR^{GJR} is more responsive than the VaR^{HV}. The VaR^{IV} also includes new volatility information quickly. The historical volatility model VaR and implied volatility VaR are distinct with respect to their levels. Because of the volatility risk premium embedded in option prices, VaR^{IV} is more negative than the alternative measures. The median VaR_t^{HV}, VaR_t^{GJR} and VaR_t^{IV} values are respectively −2.9%, −2.8% and −3.0%.

The IV adjustments provide lower absolute VaR values. The medians of $VaR_t^{IV}(NPA)$ and VaR^{IV} (PA) are −2.7% and −2.7%, indicating that the adjustments correct for the volatility risk premium efficiently. However, the extra time series information that is used for the non-parametric adjustment causes the VaR^{IV} (NPA) to be smoothed excessively and to lag significantly. The VaR^{IV} (PA) is more responsive but exhibits several extreme predictions. Although the non-parametric approach, relying on a median, is arbitrary, it does not suffer from the uncertainty in the parameter estimates.

Next, we evaluate and back-test the competing VaR measures paying particular attention to the relative performances of the time series historical volatility model approaches and the IV approaches. The VaR results are assessed using four complementary tests: the Kupiec test (1995), which measures whether the number of exceptions is consistent with the confidence level, the dynamic quantile test (Engle & Manganelli, 2004), the loss function test (Chen & Gerlach, 2013), and the conditional predictive ability test (Giacomini & White, 2006).

4.1. Unconditional test

Table 2 reports the percentage of violations (Vrate), the empirical percentage of violations divided by the theoretical percentage of violations (Vrate ratio) and the Kupiec LR statistic for the two VaR levels. The violation rate ratio should be as close as possible to one. A bold LR statistic indicates that the VaR model fails the unconditional test at the 5% significance level.

Both the violation rate and the scaled violation rate show that VaR^{IV} is violated noticeably too rarely for all markets and both VaR confidence levels. The LR test confirms statistically that the VaR^{IV} model is excessively conservative. The adjustments performed are successful in providing a more accurate VaR level. Despite the volatility risk premium corrections, the IV VaRs do not outperform the time series based VaRs. In fact, VaR^{GJR} appears to be the best VaR model of all.

In line with the results of Corrado and Miller (2005), the VaR^{IV} model is the least biased for the Nasdaq 100 market. They find that VXN is the volatility index with the least biased volatility forecast. This suggests that the higher the volatility risk premium is in a given market, the more biased the IV based VaR for that market will be.

The non-parametric adjustment provides a better VaR than the parametric adjustment. The null hypothesis of the LR test for $VaR^{IV}(NPAIV)$ is only rejected at the 5% significance level on one occasion, S&P 500%–99% VaR. On the other hand, the null is rejected at the 5% significance level four times for $VaR^{IV}(PAIV)$. Thus, the non-parametric approach is preferred because of the poor out-of-sample forecasting performance of the AR(1) model.

The unconditional test shows that historical volatility VaR models and IV adjusted VaR models provide very similar performances. In fact, VaR^{GJR} appears to be the best VaR model of all. Further tests will help to support this claim.

Table 2
Time series and implied volatilities based value-at-risk measures' unconditional performances.

	95% VaR					99% VaR				
	HV	GJR	IV	NPAIV	PAIV	HV	GJR	IV	NPAIV	PAIV
S&P										
Vrate	5.43%	5.24%	3.73%	5.49%	5.90%	1.28%	1.14%	0.61%	1.31%	1.40%
Vrate ratio	1.09	1.05	0.75	1.10	1.18	1.28	1.14	0.61	1.31	1.40
Kupiec	1.96	0.59	**19.05**	2.50	**8.28**	3.57	0.93	**9.21**	**4.62**	**7.75**
DJIA										
Vrate	5.42%	5.13%	3.98%	5.42%	5.58%	1.24%	1.21%	0.64%	1.15%	1.28%
Vrate ratio	1.08	1.03	0.80	1.08	1.16	1.24	1.21	0.64	1.15	1.28
Kupiec	1.12	0.11	**7.32**	1.12	2.12	1.74	1.32	**4.78**	0.66	2.20
NASDAQ										
Vrate	5.39%	5.65%	4.44%	5.83%	6.13%	1.22%	1.17%	0.83%	1.39%	1.65%
Vrate ratio	1.08	1.13	0.89	1.17	1.23	1.22	1.17	0.83	1.39	1.65
Kupiec	0.73	1.99	1.59	3.16	**5.81**	1.03	0.70	0.74	3.18	**8.27**

Note: The table displays the performances of five value-at-risk measures, the historical volatility (HV), GJR-GARCH (GJR), implied volatility (IV), non-parametric adjusted implied volatility (NPAIV), and parametric adjusted implied volatility (PAIV), in terms of the empirical violation rate (Vrate), the empirical violation rate over the theoretical violation rate (Vrate ratio) and the Kupiec likelihood ratio. The bold numbers indicate that the null hypothesis of the two-tailed Kupiec is rejected at the 5% significance level. The results are presented for two different VaR confidence levels (95% and 99%) and for three different indices, S&P 500, DJIA and the Nasdaq 100.

4.2. Dynamic quantile test

The dynamic quantile test (DQ) proposed by Engle and Manganelli (2004) allows us to assess the onditional performances of VaR models. This test investigates whether the violation rate is independent of certain conditions, such as past violations or volatility regimes. For the DQ test, the sequence of VaR_t^x is transformed into a hit variable as follows:

$$Hit_t\,(x) = I[r_t < VaR_t^x] - \alpha, \tag{9}$$

where the indicator variable I takes the value of one when the VaR_t^x is violated. This implies that $Hit_t\,(x)$ takes the value of one minus the VaR confidence level when a violation occurs, and minus the VaR confidence level when no violation occurs. For a correctly specified VaR, the expectation of $Hit_t\,(x)$ is zero. In a linear regression framework, $Hit_t\,(x)$ is regressed on its lags, and on any other relevant lagged variables, to test for independence. The probability of a VaR model being violated should remain unchanged under any condition, and hence, all coefficients in the regression should be equal to zero. For this purpose, Engle and Manganelli (2004) propose the conditional coverage test statistic

$$DQ_{oos} = \frac{\hat{B}'X'X\hat{B}}{\alpha(1-\alpha)}. \tag{10}$$

This statistic is $\chi^2(K)$ distributed. $\hat{B}$ is the $K \times 1$ column vector of the estimated coefficients and X is a $N \times K$ matrix, where N represents the number of observations in the regression (see Table 3).

The lagged hit variable tests for independence. The lagged VaR and squared return verify whether the violation rate is constant across different volatility regimes. The lagged change in IV identifies whether violations are affected by sudden changes in the volatility.

The VaR^{GJR} model passes the conditional efficiency tests successfully four times out of six across markets and under alternative VaR confidence levels, at the 5% significance level. The other VaRs, including VaR^{IV} and $VaR^{IV}(NPAIV)$, almost always fail the test. When excluding the VaR^{GJR} model, the null hypothesis is rejected 21 times out of 24, at the 5% significance level. The violation rate of the VaR^{GJR} model is affected the least by market conditions.

The significance of the parameter estimates indicates the relative strength or weakness of each VaR model. The effect of the volatility risk premium also appears in the regression results. The parameter estimates of the constants indicate a tendency for $\mathrm{VaR}^{\mathrm{IV}}$ to overestimate the risk, while the other VaRs underestimate it. At least one of the lagged dependent variable coefficients is significant at the 5% level in 27 cases and at the 10% level in three cases across markets and VaR confidence levels, indicating that the violations are clustered for every VaR model. Although the IV based VaRs are forward-looking and theoretically more responsive to new information, they perform no better than their competitors on this criterion.

4.3. Loss function

The two previous tests assess the performances of the VaR models individually. This highlights certain differences across models, but does not allow for a formal evaluation of the relative performances of two models. We compare the performances of IV-based and historical-based volatility VaR models directly by adopting a loss function evaluation approach. Chen and Gerlach (2013) propose a loss function for VaR models, inspired by the quantile regression loss function of Koenker and Bassett (1978). The loss function is:

$$LF = \sum_{t=1}^{N} (r_t - VaR_t)\,(\alpha - I_t), \tag{11}$$

where r_t is the return at time t, VaR_t is the forecast of a given VaR model for time t and I_t is an indicator function that takes a value of 1 if the VaR_t is violated at time t, i.e., $r_t < VaR_t$, and 0 otherwise.

Table 3
Dynamic quantile tests.

	95% VaR					99% VaR				
	HV	GJR	IV	NPAIV	PAIV	HV	GJR	IV	NPAIV	PAIV
S&P										
Constant	0.02	0.01	−0.01	0.01	0.02	0.01	0.01	−0.01	0.01	0.01
	2.86	*2.00*	*−1.57*	*2.26*	*2.90*	*3.50*	*2.19*	*−2.06*	*3.17*	*2.37*
Lag hit	0.01	−0.01	−0.01	−0.02	−0.01	−0.01	0.01	0.00	0.00	0.00
	0.29	*−0.48*	*−0.47*	*−1.13*	*−0.49*	*−0.31*	*0.34*	*−0.11*	*0.10*	*0.15*
Lag 2 hit	0.06	0.03	0.06	0.06	0.05	0.02	0.03	0.06	0.03	0.07
	3.40	*1.67*	*2.75*	*3.18*	*2.90*	*0.88*	*1.10*	*1.45*	*1.34*	*2.14*
Lag 3 hit	0.01	−0.01	0.00	0.03	0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	0.45	*−1.02*	*−0.25*	*1.81*	*0.85*	*−7.60*	*−6.12*	*−3.18*	*−6.14*	*−4.48*
Lag 4 hit	0.02	0.00	0.03	0.02	0.02	−0.01	0.01	−0.01	0.03	0.05
	0.96	*0.11*	*1.52*	*1.45*	*1.34*	*−7.67*	*0.45*	*−1.92*	*1.29*	*1.96*
Lag VaR	0.76	0.39	0.02	0.49	0.58	0.30	0.18	−0.14	0.24	0.19
	2.28	*1.45*	*0.05*	*1.47*	*1.64*	*3.27*	*1.85*	*−1.04*	*2.63*	*1.37*
Lag squared return	−0.53	−0.56	−0.31	−0.44	−0.51	−0.22	−0.11	−0.05	−0.11	−0.03
	−2.19	*−2.38*	*−1.42*	*−1.73*	*−1.81*	*−2.07*	*−1.32*	*−0.69*	*−1.04*	*−0.27*
Lag change Vix	−8.41	−14.40	−8.40	−9.79	−6.90	1.32	−1.18	−4.20	−1.86	−1.29
	−1.90	*−2.82*	*−2.52*	*−2.71*	*−1.78*	*0.64*	*−0.49*	*−1.76*	*−0.86*	*−0.45*
Adj_R^2	1%	0%	1%	1%	0%	0%	0%	1%	0%	1%
DQoos *p*-value	**0**%	**1**%	**0**%	**0**%	**0**%	**0**%	15%	**1**%	**0**%	**0**%
DJIA										
Constant	0.02	0.01	−0.01	0.02	0.02	0.01	0.01	0.00	0.01	0.01
	2.99	*0.75*	*−1.49*	*1.94*	*2.13*	*2.71*	*2.31*	*−0.42*	*1.70*	*2.26*
Lag hit	0.03	0.02	0.05	0.04	0.04	0.06	−0.01	0.00	−0.01	0.02
	1.47	*1.26*	*1.80*	*1.62*	*1.73*	*1.59*	*−2.49*	*−0.64*	*−1.19*	*0.68*
Lag 2 hit	0.05	0.02	0.04	0.03	0.04	0.01	−0.01	0.09	0.04	0.09
	2.15	*0.85*	*1.47*	*1.45*	*1.99*	*0.36*	*−5.43*	*1.13*	*0.89*	*1.74*
Lag 3 hit	0.00	−0.04	0.00	0.02	0.03	−0.01	−0.01	0.00	−0.01	−0.01
	−0.25	*−4.22*	*−0.12*	*1.17*	*1.41*	*−3.99*	*−5.55*	*−3.08*	*−4.41*	*−2.99*
Lag 4 hit	0.02	0.03	0.05	0.04	0.05	−0.01	−0.01	0.04	0.01	0.00
	0.70	*1.23*	*2.04*	*1.74*	*2.04*	*−5.00*	*−5.76*	*1.22*	*0.65*	*0.27*
Lag VaR	0.92	0.14	−0.25	0.63	0.70	0.27	0.24	0.07	0.16	0.23
	2.84	*0.35*	*−0.69*	*1.48*	*1.60*	*2.74*	*2.19*	*0.75*	*1.73*	*2.37*
Lag squared return	0.05	0.37	0.49	0.28	0.24	−0.04	0.08	0.05	0.19	0.11
	0.19	*1.41*	*1.75*	*0.89*	*0.66*	*−0.36*	*0.63*	*0.51*	*1.27*	*0.82*
Lag change IV	−6.29	−14.57	−16.24	−8.95	−8.28	−1.20	1.84	0.17	2.02	0.52
	−1.17	*−2.65*	*−3.15*	*−1.43*	*−1.44*	*−0.46*	*0.72*	*0.09*	*0.75*	*0.24*
Adj_R^2	1%	0%	1%	0%	1%	1%	0%	1%	0%	1%
DQoos *p*-value	**0**%	10%	**0**%	**0**%	**0**%	**0**%	35%	**0**%	9%	**0**%
NASDAQ										
Constant	0.03	0.02	0.00	0.03	0.03	0.01	0.01	0.00	0.01	0.01
	2.79	*2.01*	*−0.26*	*2.49*	*2.63*	*2.34*	*1.88*	*0.09*	*1.73*	*2.52*
Lag hit	−0.05	−0.03	0.00	−0.03	−0.04	−0.01	−0.01	0.00	0.00	−0.02
	−2.16	*−1.36*	*−0.04*	*−1.45*	*−1.70*	*−2.91*	*−2.56*	*−0.06*	*−0.82*	*−1.83*
Lag 2 hit	0.07	0.05	0.07	0.06	0.06	0.02	−0.01	0.04	0.05	0.09
	2.48	*2.04*	*2.20*	*2.25*	*2.21*	*0.69*	*−4.78*	*0.93*	*1.23*	*1.69*
Lag 3 hit	0.03	0.00	0.00	0.00	0.00	−0.01	−0.01	−0.01	−0.01	0.01
	1.24	*0.15*	*−0.11*	*−0.09*	*0.01*	*−4.90*	*−4.89*	*−3.12*	*−4.55*	*0.56*
Lag 4 hit	0.03	0.03	0.03	0.01	0.03	−0.01	−0.01	−0.01	0.05	0.03
	1.03	*1.21*	*1.06*	*0.52*	*1.04*	*−5.11*	*−4.80*	*−1.80*	*1.15*	*0.94*
Lag VaR	1.05	0.56	0.07	0.88	1.08	0.22	0.22	0.05	0.14	0.23
	2.91	*1.54*	*0.19*	*1.96*	*2.07*	*2.30*	*1.74*	*0.39*	*1.03*	*1.72*
Lag squared return	−0.39	−0.34	0.27	−0.05	0.02	−0.06	−0.07	0.11	0.12	0.04
	−1.56	*−1.15*	*0.93*	*−0.17*	*0.06*	*−0.55*	*−0.60*	*1.15*	*0.97*	*0.19*
Lag change IV	2.27	−2.97	−4.95	2.06	8.63	0.36	0.22	−1.61	−2.66	0.68
	0.54	*−0.54*	*−1.27*	*0.46*	*1.61*	*0.23*	*0.13*	*−0.80*	*−1.17*	*0.30*
Adj_R^2	1%	0%	1%	1%	1%	0%	0%	0%	0%	1%
DQoos *p*-value	**0**%	**3**%	**3**%	**1**%	**0**%	36%	67%	64%	**0**%	**0**%

Table 4
Loss function and loss function difference tests.

	95% VaR						99% VaR				
S&P											
Loss function											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
	6.38	6.21	6.19	6.20	6.32		1.90	1.76	1.78	1.80	1.88
Loss function differences											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		**0.17**	**0.19**	**0.18**	0.06	HV		**0.14**	**0.12**	**0.11**	0.06
GJR	**−0.17**	0.01	0.00	−0.12		GJR	**−0.14**		−0.03	−0.04	−0.09
IV	**−0.19**	−0.01	−0.01	**−0.13**		IV	**−0.12**	0.03		−0.01	−0.06
NPAIV	**−0.18**	0.00	0.01	**−0.12**		NPAIV	**−0.11**	0.04	0.01		**−0.05**
PAIV	−0.06	0.12	**0.13**	**0.12**		PAIV	−0.06	0.09	0.06	**0.05**	
DJIA											
Loss function											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
	3.96	3.81	3.84	3.85	3.90		1.18	1.08	1.08	1.09	1.12
Loss function differences											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		**0.14**	**0.12**	**0.10**	0.02	HV		**0.09**	**0.10**	**0.08**	0.06
GJR	**−0.14**		−0.02	−0.04	**−0.12**	GJR	**−0.09**		0.01	−0.01	−0.03
IV	**−0.12**	0.02		−0.02	**−0.10**	IV	**−0.10**	−0.01		−0.02	−0.04
NPAIV	**−0.10**	0.04	0.02		**−0.08**	NPAIV	**−0.08**	0.01	0.02		−0.02
PAIV	−0.02	**0.12**	**0.10**	**0.08**		PAIV	−0.06	0.03	0.04	0.02	
NASDAQ											
Loss function											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
	3.45	3.37	3.34	3.38	3.47		0.98	0.92	0.93	0.94	1.00
Loss function differences											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		**0.08**	**0.10**	0.06	−0.02	HV		**0.06**	**0.05**	0.04	−0.02
GJR	**−0.08**		0.02	−0.02	**−0.10**	GJR	**−0.06**		−0.01	−0.01	**−0.08**
IV	**−0.10**	−0.02		−0.04	**−0.13**	IV	**−0.05**	0.01		−0.01	**−0.07**
NPAIV	−0.06	0.02	0.04		**−0.09**	NPAIV	−0.04	0.01	0.01		**−0.06**
PAIV	0.02	**0.10**	**0.13**	**0.09**		PAIV	0.02	**0.08**	**0.07**	**0.06**	

Note: The table displays the values of the loss functions for the five VaR measures over three indices. The smaller the loss function, the better the VaR measure. We compare the loss function across models by computing the difference between the loss functions provided by the model in the column and that in the row. A block bootstrap is performed in order to test for the statistical significance of the loss function difference. The block length is chosen based on the following rule of thumb: $(number\ of\ observations)^{1/3}$. We rely on 1000 bootstraps.

Drawing inference based on an absolute loss function comparison is not appropriate. Chen and Gerlach (2013) propose to block bootstrap the loss function in order to obtain the distribution of this statistic and test whether the loss functions of two VaR models are statistically different.

We block bootstrap the difference between two VaR models' loss functions to test for the significance of the loss function difference, using 1000 replications with a block length of $N^{(\frac{1}{3})}$.[9] Table 4 reports the loss functions for each VaR model, VaR level and index. The loss function pairwise difference across models is presented, with statistically significant differences, according to the bootstrapped distribution with a 95% confidence interval, indicated in bold. A negative figure means that the loss function of the model on the horizontal line is lower than, and therefore preferable to, the model on the vertical line.

Despite the change in the evaluation criteria, the same conclusion is reached again: VaR^{GJR} dominantly minimizes the loss function. In a statistical sense, the historical volatility model is not outperformed by IV information for the purpose of VaR forecasting. At the 5% significance level, VaR^{GJR}, VaR^{IV} and VaR^{IV} (NPA) do not have statistically different loss functions. It is statistically difficult to distinguish competing VaRs. The uncertainty in the reported loss function is important, as is found to be the case for the uncertainty in the forecast VaR (Bams et al., 2005). The statistical test identifies VaR^{HV} and VaR^{IV} (PA) as the worst alternatives, since they are always statistically outperformed at the 5% significance level. The loss function accounts for both the number of violations and the magnitude of the difference between the VaR and realized returns. The second factor favours the VaR^{IV} and penalizes the $VaR^{IV}(NPA)$ heavily. Although VaR^{IV} is violated too rarely, the differences between r_t and VaR_t^{IV} are small. In contrast, the extreme differences between r_t and $VaR_t^{IV}(NPA)$, caused

9 The block length differs between indices in order to take into account the differences in sample size.

by the unstable $VaR_t^{IV}(NPA)$ forecasts, explain the underperformance of the $VaR^{IV}(NPA)$.

4.4. Conditional predictive ability test

The loss function bootstrapping exercise highlights the statistical underperformance of the VaR^{HV} and VaR^{IV} (PA), and also shows that the other models' unconditional performances could not be distinguished. We compare the VaR models' performances conditionally by relying on the VaR-based comparative predictive ability test of Giacomini and White (2006). This formal statistical test, which accounts for estimation uncertainty, allows us to identify the superiority of different models under different regimes.

For the application of the test, performance is defined as a loss function similar to Eq. (11):

$$L_t = (r_t - VaR_t)(\alpha - I_t). \tag{12}$$

As was explained by Giacomini and Komunjer (2005), this "tick function" is particularly suitable for the assessment of quantile predictions. Furthermore, it is comparable to the loss function used in the bootstrapping exercise, allowing for a comparison of the two tests.

Under the null hypothesis, the loss difference between competing models follows a martingale difference sequence.[10] The assimilated Wald-type test statistic is:

$$CPA = T\left(T^{-1}\sum_{t=1}^{T-1} I_t \triangle L_{t+1}\right)' \widehat{\Omega}' \left(T^{-1}\sum_{t=1}^{T-1} I_t \triangle L_{t+1}\right), \tag{13}$$

where T is the number of observations, I_t is the conditioning instrument at time t, ΔL_t is the loss difference at time t, and $\widehat{\Omega}$ is the Newey-West estimate of the variance–covariance of $I_t \triangle L_{t+1}$. We rely on the same conditioning instruments $(1, \triangle L_t)$ as Giacomini and White (2006) and Santos et al. (2013). The statistic is $\chi^2(T)$ distributed, and Table 5 reports the CPA associated p-values across models. A negative figure means that the loss function of the model on the horizontal line is lower than, and therefore preferable to, the model on the vertical line.

The CPA results follow a pattern similar to the bootstrapping test results. The VaR^{HV} and VaR^{IV} (PA) modelsremain the worst candidates. The VaR^{HV} model is statistically outperformed in 19 of the 24 pair-wise comparisons at the 5% significance level, while the VaR^{IV} (PA) model is outperformed in 15. In contrast to the bootstrapping results, the CPA test rejects the null hypothesis for a few more pairs of models. The conditional nature of the test provides some evidence that the 'best model' might depend on the market regime. However, the rejection of the new null hypothesis appears less systematic across markets and confidence levels. The estimation uncertainty captured by the test does not suggest a statistical difference in the performances of competing models. Furthermore, the conditional setting also confirms that an IV-based model does not outperform a historical volatility model.

10 Hence, the predictive ability is equal and unpredictable.

4.5. Robustness check: ARMA(1,1)

The various tests highlight the poor performance of the VaR^{IV} (PA). The model is outperformed statistically by all other candidates. The performance of the VaR^{IV} (PA) is related directly to its ability to forecast the volatility risk premium adequately. The choice of an AR(1) specification for modelling the volatility risk premium could be inappropriate and might explain the underperformance of the VaR^{IV} (PA) model. In order to investigate this claim, an ARMA(1,1) model is used to forecast the volatility risk premium. Although the other candidates' predictions remain constant, this richer specification changes the VaR^{IV} (PA).

The results demonstrate that an ARMA(1,1) is not a better alternative to a simple AR(1) specification for adjusting for the volatility risk premium. The newly obtained VaR^{IV} (PA) model fails the majority of the unconditional tests. The Kupiec test is always rejected at the 5% significance level for both S&P 500 and Nasdaq 100 data, while, for the DJIA returns, the 99% VaR^{IV} (PA) model fails the Kupiec test at the 10% significance level. Furthermore, the model fails all conditional efficiency tests for all three markets. The dynamic quantile test is rejected systematically at the 1% significance level.

The pair-wise comparisons demonstrate that the ARMA(1,1) is inferior to the simpler AR(1). The bootstrap and CPA tests conclude that the VaR^{IV} (PA) model is outperformed statistically at the 5% significance level for 17 out of 18 comparisons across models excluding $\mathrm{VaR}^{\mathrm{HV}}$. On average, the VaR^{HV} model also minimizes the loss function compared to the VaR^{IV} (PA) model, but the tests do not provide sufficient evidence to statistically distinguish the two models' performances. The forecasts provided by the ARMA(1,1) model appear noisier than those obtained from the AR(1). The tick function penalises the magnitude of extreme forecasts heavily. We interpret these results as evidence that the volatility risk premium is generally not captured properly by linear models. Because the main conclusions remain unchanged, and for the sake of brevity, the results are not reported here, but all statistical results are available from the authors upon request.

4.6. Robustness check: alternative innovation specification

The filtered historical simulation accounts for the non-normality of innovations, and is applied systematically in this paper. This approach improves the forecasting abilities of all models considered substantially. Nevertheless, it is possible to rely on non-normal parametric innovations for the historically estimated models. This section evaluates the impacts of different innovation treatments and provides the VaR^{GJR} model based on a skewed-t distribution (Hansen, 1994). The parametric form accounts for both the fat-tails and the asymmetry of innovations. Although the other candidates' predictions remain constant, the parametric treatment of innovations changes the VaR^{GJR} model. In some instances, the parameter estimates obtained did not allow us to compute the skewed-t distribution quantile, in which case a student-t distribution was alternatively adopted.

Table 5
Conditional predictive ability test.

95% VaR						99% VaR					
S&P											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		+ **0%**	+ **1%**	+ **0%**	+ **2%**	HV		+ **0%**	+ **4%**	+ **2%**	+ **3%**
GJR	− **0%**		+ 11%	+ **0%**	− **0%**	GJR	− **0%**		− **4%**	− **2%**	− **0%**
IV	− **1%**	− 11%		− 90%	− **3%**	IV	− **4%**	+ **4%**		− 87%	− 15%
NPAIV	− **0%**	− **0%**	+ 90%		− **0%**	NPAIV	− **2%**	+ **2%**	+ 87%		− **1%**
PAIV	− **2%**	+ **0%**	+ **3%**	+ **0%**		PAIV	− **3%**	+ **0%**	+ 15%	+ **1%**	
DJIA											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		+ **0%**	+ **2%**	+ **1%**	+ 11%	HV		+ **1%**	+ **0%**	+ **0%**	+ 20%
GJR	− **0%**		− 46%	− 13%	− **1%**	GJR	− **1%**		+ 35%	− 45%	− 18%
IV	− **2%**	46%		− **3%**	− **0%**	IV	− **0%**	− 35%		− 34%	− **4%**
NPAIV	− **1%**	+ 13%	− **3%**		− **1%**	NPAIV	− **0%**	+ 45%	+ 34%		− 17%
PAIV	− 11%	+ **1%**	+ **0%**	+ **1%**		PAIV	− 20%	+ 18%	+ **4%**	+ 17%	
NASDAQ											
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		+ **4%**	+ **0%**	+ **1%**	− **3%**	HV		+ **2%**	+ **2%**	+ 13%	− 21%
GJR	− **4%**		+ 21%	− **0%**	− **1%**	GJR	− **2%**		− 31%	− 11%	+ **3%**
IV	− **0%**	− 21%		− 15%	− **0%**	IV	− **2%**	+ 31%		− **1%**	− **4%**
NPAIV	− **1%**	+ **0%**	+ 15%		− **1%**	NPAIV	− 13%	+ 11%	+ **1%**		− 7%
PAIV	+ **3%**	+ **1%**	+ **0%**	+ **1%**		PAIV	+ 21%	+ **3%**	+ **4%**	+ 7%	

Note: The table displays the sign (+, −) of the average differences in losses between each pair of models. The smaller the losses, the better the VaR measure. We compare the losses across models by computing the average loss difference between the model in the column and that in the row. The values below provide statistical evidence of the models' relative predictive abilities, indicated by the *p*-values associated with the CPA.

The good performance of the VaR^{GJR} model remains unaffected by the adoption of a parametric form of innovations. The skewed-*t* distribution captures the empirical features exhibited by empirical innovations well. The VaR^{GJR} passes the majority of the unconditional and conditional efficiency tests at the 5% significance level. The Kupiec test fails to be rejected at the 5% significance level in all six instances. In addition, the dynamic quantile test is rejected at the 5% significance level only once (S&P 500%–95% VaR).

The VaR^{GJR} model also minimizes both the loss and the tick function. The CPA tests show that the use of skewed-*t* innovations improves the 95% confidence level VaR model forecasting ability. The skewed-*t* VaR^{GJR} model outperforms its competitors significantly 11 times, at the 5% significance level. This number decreases to seven when the filtered historical simulation VaR^{GJR} is used. All unreported statistical results of the robustness checks are available from the authors upon request.

4.7. Robustness check: maturity mismatch

Our results show that IV does not outperform historical volatility information based VaR forecasts. This contradicts the previous literature on forecasting the volatility. Maturity mismatch is a factor that has been widely claimed to impair the forecasting ability of IV (Christensen & Prabhala, 1998; Yu et al., 2010). We are evaluating the models' abilities to provide accurate one-day out-of-sample VaR forecasts, while volatility indices represent the market expectation of volatility over the next 30 days, not the next day. Although the volatility term structure implied from options has not been found to be very steep (Christoffersen et al., 2013; Stein, 1989), a completely flat term structure cannot be assumed.

Table 6
Monthly value-at-risk measures: performances and tests.

	95% VaR					99% VaR				
	HV	GJR	IV	NPAIV	PAIV	HV	GJR	IV	NPAIV	PAIV
S&P										
Vrate	4.80%	4.33%	3.08%	4.86%	5.02%	1.02%	0.90%	0.59%	1.14%	1.33%
Vrate ratio	0.96	0.87	0.62	0.97	1.00	1.02	0.90	0.59	1.14	1.33
Kupiec	0.42	**4.99**	**45.67**	0.20	0.00	0.02	0.51	**10.25**	0.93	**5.18**
DQoos p-value	**0%**	**1%**	**0%**	**0%**	**0%**	12%	90%	25%	57%	**1%**
DJIA										
Vrate	5.23%	4.62%	3.79%	5.64%	5.99%	1.40%	1.34%	1.02%	1.50%	1.47%
Vrate ratio	1.05	0.92	0.76	1.13	1.20	1.40	1.34	1.02	1.50	1.47
Kupiec	0.33	0.97	**10.48**	2.61	**6.12**	**4.56**	3.28	0.01	**6.81**	**6.02**
DQoos p-value	**0%**	**3%**	**0%**	**0%**	**0%**	7%	8%	66%	**2%**	**2%**
NASDAQ										
Vrate	6.39%	4.87%	4.39%	5.70%	5.79%	1.30%	1.39%	0.91%	1.52%	1.70%
Vrate ratio	1.28	0.97	0.88	1.14	1.16	1.30	1.39	0.91	1.52	1.70
Kupiec	**8.68**	0.08	1.85	2.26	2.85	1.97	3.18	0.18	**5.46**	**9.32**
DQoos p -value	**0%**	18%	**2%**	**1%**	**3%**	58%	17%	89%	10%	**0%**

Note: The table displays the performances of five value-at-risk measures, the historical volatility (HV), GJR-GARCH (GJR), implied volatility (IV), non-parametric adjusted implied volatility (NPAIV), and parametric adjusted implied volatility (PAIV), in terms of the empirical violation rate (Vrate), the empirical violation rate over the theoretical violation rate (Vrate ratio), the Kupiec likelihood ratio and the dynamic quantile test p-values (DQoos p-value). The bold numbers indicate that the null hypothesis of the two-tailed Kupiec is rejected at the 5% significance level. The results are presented for two different VaR confidence levels (95% and 99%) and for three different indices, S&P 500, DJIA and the Nasdaq 100.

Unfortunately, this is the only maturity for which the volatility indices are available.[11] It is also not possible to compute a model-free implied volatility accurately from very short maturity options, because of the lack of liquidity for these contracts. The relative success of the VaR^{GJR} model can be explained by its ability to adapt to different horizons. We verify this hypothesis by replicating the previous analysis for one-month out-of-sample VaRs. The GJR-GARCH volatility tends to revert quickly to the long-run volatility, producing overly smooth long-term forecasts.

Furthermore, the volatility forecasting literature highlights the high sensitivity of the results to the choice of the forecasting horizon (Cheng & Fung, 2012; Ederington & Guan, 2002; Frijns et al., 2010). Although the one-month VaR is neither standard nor required by the regulator, it is a proper test of the robustness of our results (see Table 6).

The monthly tests are aligned to the one-day results of Tables 2 and 3. For the Kupiec test at the monthly horizon, we observe a convergence amongst the VaR^{HV}, VaR^{IV}, VaR^{GJR} and VaR^{IV} (NPA) model performances. Repeatedly, there is no clear evidence that IV based VaRs outperform historical volatility model VaRs. VaR^{GJR} no longer dominates the debate, but IV based VaRs still do not provide a better alternative. For the 99% confidence level, all VaR models except for the VaR^{GJR} model with the Nasdaq 100 fail the conditional efficiency test at the 5% significance level. In the majority of cases, the non-overlapping lag hit variables are significant at the 5% level. The likelihood of other VaR models being violated is largely unaffected by the remaining variables. The dynamic quantile test is rejected only twice at the 5% significance level for the VaR^{GJR} model. Similarly to the daily forecast exercise, this is the smallest number of rejections among all competing models (see Table 7).

The good performance of the VaR^{GJR} model, in term of the loss function, is robust to the change in maturity. VaR^{GJR} minimizes the loss function in 21 of the 24 pair-wise comparisons. This superiority is statistically significant at the 5% level 11 and 6 times according to the bootstrap and CPA tests, respectively.[12] Thus, it can be concluded that the relatively good performance of time series information compared to IV information in the VaR context is not explained by a maturity mismatch.

We emphasise the discrepancy between the conclusions drawn by the volatility forecasting literature and our results. Our findings show that predicting the volatility and predicting a certain quantile are two different objectives. As is the case for IV, the optimal forecast of the volatility does not necessarily translate into the optimal forecast of a quantile. Giot and Laurent (2004) reach a similar conclusion when assessing the performances of intra-day data based VaR forecasts. Two factors may explain why IV is the best source of information for predicting the volatility, but not for predicting quantiles.

The relative predictive power of volatility forecasts is sensitive to the precision of the volatility measure used. IV is found to be a good forecast for standard volatility proxies, but Martin et al. (2009) re-assess this finding using a set of noise-robust volatility measures, and conclude that IV does not outperform other models for predicting the S&P 500 volatility. Quantile prediction requires a greater degree of precision, which explains the disappointing results from using IV information for forecasting VaR.

Both the implicit (Prokopczuk & Wese Simen, 2014) and explicit (Jorion, 1995) volatility risk premium adjustments in the volatility forecasting literature rely on important assumptions. The volatility risk premium is only linearly

[11] The CBOE did recently launch a new volatility index for a shorter maturity. However, the VXST, which is the VIX equivalent for nine-day maturity options, was only launched in 2011.

[12] The CPA was adapted to account for the fact that the monthly setting requires a multi-step conditional predictive ability test.

Table 7
Monthly VaR models' relative performances.

	95% VaR						99% VaR				
S&P											
Loss function											
	28.25	26.53	27.89	27.23	27.96		7.87	7.44	7.96	7.77	8.11
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		+* **1**%	+ 53%	+ 21%	+ 45%	HV		+ 20%	− 36%	+ 91%	− 33%
GJR	−* **1**%		−* **1**%	− 6%	−* **2**%	GJR	− 20%		−* **3**%	−* 7%	−* 9%
IV	− 53%	+* **1**%		+ 10%	− 59%	IV	+ 36%	+* **3**%		+ 10%	− 35%
NPAIV	− 21%	+ 6%	− 10%		−* 14%	NPAIV	− 91%	+* 7%	− 10%		− 35%
PAIV	− 45%	+* **2**%	+ 59%	+* 14%		PAIV	+ 33%	+* 9%	+ 35%	+ 35%	
DJIA											
Loss function											
	18.78	17.94	18.22	18.01	18.25		5.26	5.07	5.05	5.10	5.19
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		+* 6%	+ 55%	+ 24%	+ 40%	HV		+ **5**%	+ 68%	+ 70%	+ 94%
GJR	−* 6%		− 51%	− 93%	− 75%	GJR	− **5**%		+ 72%	− 62%	− 78%
IV	− 55%	+ 51%		+ 64%	− 93%	IV	− 68%	− 72%		− 25%	− 36%
NPAIV	− 24%	+ 93%	− 64%		− 70%	NPAIV	− 70%	+ 62%	+ 25%		− 81%
PAIV	− 40%	+ 75%	+ 93%	+ 70%		PAIV	− 94%	+ 78%	+ 36%	+ 81%	
NASDAQ											
Loss function											
	16.58	15.67	16.53	16.20	16.44		4.60	4.28	4.27	4.25	4.37
	HV	GJR	IV	NPAIV	PAIV		HV	GJR	IV	NPAIV	PAIV
HV		+* 12%	+ 85%	+ 60%	+ 94%	HV		+ 15%	+ 30%	+ 27%	+ 72%
GJR	−* 12%		−* **0**%	−* 8%	−* 6%	GJR	− 15%		+ 39%	+ 63%	− 93%
IV	− 85%	+* **0**%		+ 33%	+ 20%	IV	− 30%	− 39%		+ 63%	− 38%
NPAIV	− 60%	+* 8%	− 33%		− 28%	NPAIV	− 27%	− 63%	− 63%		− 44%
PAIV	− 94%	+* 6%	− 20%	+ 28%		PAIV	− 72%	+ 93%	+ 38%	+ 44%	

Note: The loss function differences between model pairs are assessed statistically using the following two approaches. The first row provides the loss function difference sign (+, −) and statistical significance based on bootstrapped distributions. The second row provides the *p*-values associated with the conditional predictive ability test.

* Indicates significant differences at the 5% level.

dependent on the volatility level, and independent of returns. These assumptions are sufficient for forecasting the average volatility level over time, but our results suggest that they are not sufficient for forecasting extreme quantiles where non-linearity and return dependence matter.

Supporting this conjecture, Charoenwong et al. (2009) show that IV is unable to provide better volatility forecasts than historical volatility models during high volatility regimes. In addition, the dynamic quantile test performed in this paper demonstrates that the adjustment allows for the correction of the average level of violations, but leads to an increased probability of violations during high volatility regimes. A richer dynamic that can account for non-linear dependencies between the volatility risk premium, volatility, returns and innovations is necessary in order to adjust IV for VaR forecasting purposes.

Overall, we find IV to have limited utility for predicting either quantiles of future returns or VaR. The volatility risk premium adjustment that is proved to be performance enhancing in volatility forecasting (Prokopczuk & Wese Simen, 2014) or portfolio allocation (DeMiguel et al., 2013) is not so for VaR forecasting. This exercise underlines the difference between volatility forecasting and quantile forecasting.

5. Conclusion

This paper casts new light on the comparison between options implied volatility and historical return time series information. The efficiency of these two sources of information has been evaluated and compared on numerous occasions, based on their predictive power for the future realized volatility. We have re-evaluated the relative benefits of these two information sources with a more pragmatic objective. Do IV or historical volatility models provide the best Value-at-Risk estimate?

We also formally propose a correction of the IV for the volatility risk premium. IV estimates are upward biased and are found to lead to overstated VaR values because of the negative price of the volatility risk that is embedded in options. Our parametric and non-parametric adjustments reduce this bias and account for the presence of the volatility risk premium formally. Previous studies have mainly ignored the volatility risk premium, or accounted for it in an indirect and unconscious way.

The results for three different indices, and using different evaluation criteria, all point in the same direction. IV-based VaRs do not outperform historical volatility model VaRs. A simple VaR based on a GARCH that accounts for leverage provides very satisfactory results. IV-based VaRs cannot outperform this GARCH VaR, even after the volatility risk premium adjustment. In fact, it appears that the GARCH VaR outperforms most of its competitors on many occasions.

This result is in stark contrast to the volatility forecasting literature, where IV is shown to provide better forecasts than historical volatility models. We show that forecasting the volatility is different from forecasting a certain quantile of the return distribution, and hence, the conclusion that IV is superior to the historical volatility model cannot be translated easily into the risk management context of the VaR measure.

References

Bali, T. G., & Zhou, H. (2016). Risk, uncertainty, and expected returns. *Journal of Financial and Quantitative Analysis*, *51*(03), 707–735.

Bams, D., Lehnert, T., & Wolff, C. C. (2005). An evaluation framework for alternative VaR-models. *Journal of International Money and Finance*, *24*(6), 944–958.

Barone-Adesi, G., Bourgoin, F., & Giannopoulos, K. (1998). Market risk: don't look back. *Risk*, *11*, 100–103.

Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price variability. *Journal of Banking & Finance*, *5*(3), 363–381.

Berkowitz, J., & O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *The Journal of Finance*, *57*(3), 1093–1111.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Blair, B. J., Poon, S. H., & Taylor, S. J. (2010). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, *105*(1), 5–26.

Britten-Jones, M., & Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, *55*(2), 839–866.

Canina, L., & Figlewski, S. (1993). The informational content of implied volatility. *Review of Financial Studies*, *6*(3), 659–681.

Carr, P., & Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, *22*(3), 1311–1341.

Charoenwong, C., Jenwittayaroje, N., & Low, B. S. (2009). Who knows more about future currency volatility?. *Journal of Futures Markets*, *29*(3), 270–295.

Chen, Q., & Gerlach, R. H. (2013). The two-sided Weibull distribution and forecasting financial tail risk. *International Journal of Forecasting*, *29*(4), 527–540.

Cheng, X., & Fung, J. K. (2012). The information content of model-free implied volatility. *Journal of Futures Markets*, *32*(8), 792–806.

Chernov, M. (2007). On the role of risk premia in volatility forecasting. *Journal of Business and Economic Statistics*, *25*(4), 411–426.

Chong, J. (2004). Value at risk from econometric models and implied from currency options. *Journal of Forecasting*, *23*(8), 603–620.

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, *50*(2), 125–150.

Christoffersen, P., Heston, S., & Jacobs, K. (2013). Capturing option anomalies with a variance-dependent pricing Kernel. *Review of Financial Studies*, *26*(8), 1963–2006.

Christoffersen, P., & Mazzotta, S. (2005). The accuracy of density forecasts from foreign exchange options. *Journal of Financial Econometrics*, *3*(4), 578–605.

Corrado, C. J., & Miller Jr, T. W. (2005). The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets*, *25*(4), 339–373.

Day, T. E., & Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, *52*(1), 267–287.

DeMiguel, V., Plyakha, Y., Uppal, R., & Vilkov, G. (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis*, *48*(6), 1813–1845.

Ederington, L. H., & Guan, G. (2002). Is implied volatility an informationally efficient and effective predictor of future volatility?. *Journal of Risk*, *4*, 29–46.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*(4), 367–381.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Fleming, J. (1998). The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance*, *5*(4), 317–345.

Frijns, B., Tallau, C., & Tourani-Rad, A. (2010). The information content of implied volatility: evidence from Australia. *Journal of Futures Markets*, *30*(2), 134–155.

Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics*, *23*(4), 416–431.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Giot, P. (2005). Implied volatility indexes and daily value at risk models. *The Journal of Derivatives*, *12*(4), 54–64.

Giot, P., & Laurent, S. (2003). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics*, *18*(6), 641–663.

Giot, P., & Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, *11*(3), 379–398.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*, 705–730.

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

Jeon, J., & Taylor, J. W. (2013). Using CAViaR models with implied volatility for value-at-risk estimation. *Journal of Forecasting*, *32*(1), 62–74.

Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *Review of Financial Studies*, *18*(4), 1305–1342.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *The Journal of Finance*, *50*(2), 507–528.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, *46*, 33–50.

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*(1), 53–89.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*(2), 73–84.

Lamoureux, C. G., & Lastrapes, W. D. (1993). Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *Review of Financial Studies*, *6*(2), 293–326.

Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, *24*(11), 1005–1028.

Martin, G. M., Reidy, A., & Wright, J. (2009). Does the option market produce superior forecasts of noise-corrected volatility measures? *Journal of Applied Econometrics*, *24*(1), 77–104.

Mittnik, S., & Paolella, M. S. (2000). Conditional density and value-at-risk prediction of Asian currency exchange rates. *Journal of Forecasting*, *19*(4), 313–333.

Mixon, S. (2009). Option markets and implied volatility: Past versus present. *Journal of Financial Economics*, *94*(2), 171–191.

Nieto, M. R., & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, *32*(2), 475–501.

Pong, S., Shackleton, M. B., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR (FI) MA models. *Journal of Banking & Finance*, *28*(10), 2541–2563.

Prokopczuk, M., & Wese Simen, C. (2014). The importance of the volatility risk premium for volatility forecasting. *Journal of Banking & Finance*, *40*, 303–320.

Rosenberg, J. V., & Engle, R. F. (2002). Empirical pricing Kernels. *Journal of Financial Economics*, *64*(3), 341–372.

Santos, A. A., Nogales, F. J., & Ruiz, E. (2013). Comparing univariate and multivariate models to forecast portfolio value-at-risk. *Journal of Financial Econometrics*, *11*(2), 400–441.

Stein, J. (1989). Overreactions in the options market. *The Journal of Finance*, *44*(4), 1011–1023.

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

Taylor, S. J. (2005). *Asset price dynamics, volatility, and prediction*. Princeton University Press.

Taylor, S. J., & Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, *4*(4), 317–340.

Taylor, S. J., Yadav, P. K., & Zhang, Y. (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance*, *34*(4), 871–881.

Tsiaras, L. (2009). The forecast performance of competing implied volatility measures: The case of individual stocks (No. F-2009-02). University of Aarhus, Aarhus School of Business, Department of Business Studies.

Yu, W. W., Lui, E. C., & Wang, J. W. (2010). The predictive power of the implied volatility of options traded OTC and on exchanges. *Journal of Banking & Finance*, *34*(1), 1–11.