The Information Content of the VDAX Volatility Index and Backtesting Daily Value-at-Risk Models

Ihsan U. Badshah
Auckland University of Technology, Department of Finance,
Private Bag 92006, 1142 Auckland, New Zealand.
Phone: +64-9-9219999 ext. 5394, Fax: +64-9-9219940,
Email: ihsan.badshah@aut.ac.nz

Abstract

This paper examines the information content of the new VDAX volatility index to forecast daily value-at-risk (VaR) estimates and compares its VaR forecasts with the VaR forecasts of the filtered historical simulation (FHS) and RiskMetrics models. The daily VaR models were backtested from January 1992 through May 2009 using unconditional coverage, independence, and conditional coverage tests. A quadratic score was also estimated for each of the models. We found that the information content of implied volatility was superior to that of the historical volatility for the daily VaR forecasts of a portfolio of the DAX 30 stock index: implied volatility (VDAX) and combined (implied volatility plus GJR) VaR models outperformed the VaR models of the FHS (GJR) and RiskMetrics. Finally, the quadratic score also supports the use of implied volatility VaR models. Our findings have implications for traders, risk managers, and regulators.

Keywords: backtesting, filtered historical simulation, implied volatility, value at risk, VDAX

JEL Classifications: G13, C52, C53
1 Introduction

Volatility forecasting is important to both financial practitioners and academics. This is very much true in option pricing, option trading, hedging derivative positions, risk management activities, stock selection and portfolio diversification, all of which require accurate volatility modeling and forecasting. Fortunately, we have seen tremendous development in volatility forecasting models since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model by Engle (1982). Since then, many ARCH models have been developed that attempt to forecast volatility using historical information. In 1986, Bollerslev (1986) extended the ARCH model proposed by Engle (1982), to a generalized ARCH model called GARCH. Many extensions have been made to the GARCH model (see Poon and Granger, 2003; Poon and Granger, 2005); for example, Glosten, Jagannathan and Runkle (hereafter GJR)(1993) proposed the GJR model, an extension of the GARCH model that is used to account for asymmetry. However, the growing literature supports using implied volatility (IV) instead, calling IV the best forecast of future realized volatility (RV); they found that the information content of IV is richer and superior to that of historical volatility (HV) when forecasting the future RV of the underlying asset (e.g., Day and Lewis, 1992; Christensen and Prabhala, 1998; Fleming, 1998; Dumas, Fleming and Whaley, 1998; Blair et al., 2001; Ederington and Guan, 2002; Poon and Granger, 2003; Mayhew and Stivers, 2003; and Martens and Zein, 2004).

IV can be recovered by inverting the Black-Scholes (1973) formula. However, Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) derived model-free implied volatility (MFIV) under the assumptions of pure diffusion and asset price processes with jumps. They showed that the information content of MFIV is richer and superior to that of Black-Scholes implied volatility (BSIV): the MFIV measure accounts for all available strike prices, whereas the BSIV measure is a point-based IV, where each strike price has a separate IV. Additionally, BSIV is subject to both model risk and market efficiency, while MFIV is only subject to market efficiency (see Poon and Granger, 2003).

Deutsche Börse has launched a new VDAX volatility index for the German DAX 30 stock index that incorporates more robust information on options by using MFIV measures. In addition, the new VDAX measure aligns with the consensus view of option traders (who are usually professionals) about the future direction of the volatility of the stock market over the next 30 days. The rationale for adopting MFIV measures was to account for both out-of-the-

---

1 Chicago Board Options Exchange uses a similar MFIV measure for the VIX and VXN, which are derived from S&P 500 and NASDAQ options, respectively.
money (OTM) put and call options (i.e., volatility skew). The new VDAX measure incorporates both types of options; therefore, it moves with changes in option prices. For example, a negative or positive shock to the market induces adjustments in hedging and trading strategies; this consequently triggers changes in the prices of one type of option (i.e., put or call). It then moves in the direction of the market demand for particular types of options and underlying stocks (see, e.g., Bollen and Whaley, 2004). In addition, Liu et al. (2005) found that the rare-event premiums play an important role in generating the volatility skew patterns observed for options across moneyness and that these rare events are embedded in the OTM options. Camara and Heston (2008) derived an option pricing model that accounted for both OTM put and call options. They derived the extreme negative events from OTM puts and the extreme positive events from OTM calls. Thus, The new VDAX index should be an excellent instrument for forecasting volatility (or VaR) for the underlying DAX 30 portfolio as it is implied by OTM options, which are informed on future events and contain information on negative and positive jumps (see, e.g., Pan, 2002; Liu et al., 2005; Camara and Heston, 2008; Bates, 2008).

There are some related studies that use the information content of IV in their volatility forecasting models and find that IV subsumes almost all information for forecasting future RV. For example, Blair et al. (2001) studied the information content of the VIX (now VXO) and intraday returns for the S&P 100 stock index and then compared them for the sample period from 1987 to 1999. Their in-sample forecasting showed that all relevant information is provided by the IV (i.e., implied from the VIX) and that there is not much incremental information in intraday index returns. However, for out-of-sample forecasts, the VIX provides superior and accurate forecasts for all forecast performance measures and horizons (i.e., from 1 to 20 days); moreover, the incremental forecasting information in the intraday returns is insignificant. Ederington and Guan (2002) studied the importance of IV forecasts while using S&P 500 future options, finding that IV has superior forecasting power and subsumes the information in HV. Similarly, Martens and Zein (2004) confirmed that the GARCH model extended with IV provides better forecasts than the GARCH model extended with RV (obtained from high-frequency intraday returns). Finally, the study by Giot (2005), which is slightly different from the above two studies, assesses the information content of IV indices such as the VIX (now the VXO) and the VXN in a daily VaR framework while studying time periods that include both bull and bear markets. The performances of the VaR

2 Whaley (2000) referred to the volatility index as an “investors’ fear gauge”.
3 Similarly, Pan (2002) showed that volatility skew is primarily due to investors’ fear of large adverse jumps.
models were evaluated using unconditional coverage, independence and conditional coverage tests. Backtesting results showed that IV indices provide superior VaR forecasts and thus fewer VaR violations over time relative to the time-series-based VaR forecasts. Moreover, the performance was stable during market turmoil.

This paper contributes to the literature on market risk management, volatility modeling, and forecasting. The field of volatility modeling and forecasting, especially when predicting future RV by using IV, HV or a combination of the two, represents an interesting and less trite area of research in finance. In this paper, we incorporated the information content of the new VDAX volatility index into daily VaR models, and the resulting VaR forecasts were compared with the VaR forecasts from the FHS model (GJR) and the RiskMetrics model at confidence levels of 99%, 97.5% and 95%. The daily VaR models were then backtested using unconditional coverage, independence, and conditional coverage tests; furthermore, a quadratic score was estimated for each of the VaR models for the period from January 1, 1992 through May 29, 2009. However, our paper differs from the above-cited studies in several ways. First, they focus on using the BSIV index to forecast future RV for U.S. stock markets, whereas we focused on using a robust MFIV volatility index to forecast daily VaR for the German stock market. Second, we compared the implied daily VaR forecasts with the forecasts from the famous approach by Barone-Adesi et al. (1998), the filtered historical simulation; to our knowledge, no other study has ever compared it with implied VaR forecasts. Finally, we used much more detailed and robust backtesting techniques on the very long data series to evaluate different VaR models for daily market risk management.

Our main finding was that the MFIV (the new VDAX) index subsumed almost all information required for actual volatility in VaR models. The backtesting results show that the new VDAX index contained significant information regarding volatility in VaR models and that the number of implied VaR violations was not significantly different from the set coverage rates, thereby yielding fewer VaR exceptions and clusters of exceptions. Consequently, the null hypotheses regarding independence and conditional coverage tests were never rejected for implied volatility, implied volatility plus GJR, and FHS VaR models. However, the null hypotheses for the RiskMetrics model were rejected at lower confidence levels. Finally, the quadratic scores also favored implied VaR models.

This paper is organized as follows. In Section 2, we discuss the new VDAX index and its construction. In Section 3, we discuss the data. In Section 4, we specify daily VaR models. In Section 5, we discuss backtesting techniques. The empirical results are presented in Section 6. In Section 7, we summarize and conclude.
2 The New VDAX Volatility Index

The Deutsche Börse and Goldman Sachs jointly developed the methodology for the new VDAX index. It is based on options on the DAX 30 index. Like VIX the VDAX measure is also based on MFIV measure. Hence, it accounts for IVs across all options of a given time to expiry (it accounts for volatility skew). In fact, the main VDAX index is further based on eight constituents volatility indices that expire in 1, 2, 3, 6, 9, 12, 18, and 24 months, respectively. The main VDAX is designed as a rolling index at a fixed 30 days to expiry via a linear interpolation of the two nearest of the eight available sub-indices. The price history for the VDAX is calculated back to the years 1992. The VDAX and its eight sub-indices are updated every minute. The sub-indices are calculated according to the formula below:

\[ VDAX_i = 100 \cdot \sqrt{\sigma_i^2}, \]  
\[ \sigma_i^2 = \frac{2}{T_i} \sum_j \frac{\Delta K_{ij}}{K_{ij}^2} \cdot R_i \cdot M(K_{ij}) \cdot \frac{1}{T_j} \left( \frac{F_i}{K_{i,0}} - 1 \right)^2, \quad i = 1, 2, ..., 8. \]  

Where \( T_i \) is time to expiry of the \( i \)th DAX 30 option (best bid and best ask of all DAX 30 options), \( F_i \) is forward price derived from the prices of the \( i \)th DAX 30 options, for which the absolute difference between call and put prices is the smallest as \( F_i = K_{\min[C-P]} + R_i \cdot (C - P) \), \( K_{ij} \) the exercise price of the OTM option of the \( i \)th DAX 30 option expiry month, \( \Delta K_{ij} = \frac{K_{ij+1} - K_{ij-1}}{2} \) is the interval between the strike prices, \( K_{i,0} \) the highest exercise price below forward price, \( R_i \) is the refining factor equal to \( R_i = e^{\delta \cdot T_i} \), \( r \) is the risk free rate, \( M(K_{ij}) \) is the price of the option \( K_{ij} \), whereby \( K_{ij} \neq K_{i,0} \), and \( M(K_{i,0}) \) is the average of the put and call prices at exercise price \( K_{i,0} \).

Finally, the main VDAX is designed as a rolling index at a fixed 30 days to expiry via a linear interpolation of the two nearest of the eight available sub-indices as follows.

\[ VDAX = \sqrt{\left( T_i \cdot VDAX_i^2 \cdot \frac{N_{T_{i+1}} - N_T}{N_{T_{i+1}} - N_{T_i}} + T_{i+1} \cdot VDAX_{i+1}^2 \cdot \frac{N_T - N_{T_i}}{N_{T_{i+1}} - N_{T_i}} \right) \cdot \frac{N_{365}}{N_T}}, \]  

where \( N_T \) time to expiry of the \( i \)th DAX 30 index option, \( N_{T_{i+1}} \) time to expiry of the \( i+1 \)th option, \( N_T \) time for next days, and \( N_{365} \) time for standard year.\(^4\)

\(^4\) See, for further detail on the VDAX construction, the Deutsche Börse and Goldman Sachs methodology for
3 Data

This paper employs data from two sources. We obtained the daily time-series price data for the DAX 30 stock index from Financial DataStream. The data on the new VDAX were obtained from the Deutsche-Börse. The daily data for both the DAX 30 and the VDAX indices cover a period of 17 years and 5 months, from January 1, 1992, to May 29, 2009, for a total of 4,543 trading days.

![DAX 30 stock index returns](image1)

![VDAX volatility index level](image2)

**Figure 1.** DAX 30 stock index returns (%) and the VDAX volatility index level (%) from January 1992 through May 2009.

VSTOXX on the STOXX website: [www.stoxx.com](http://www.stoxx.com), the same methodology is also used for the construction of the VDAX volatility index.
Figure 1 shows the daily continuously compounded returns (%) of the DAX 30 stock index and the daily closing level (%) of the VDAX volatility index from January 1992 through May 2009. As can be seen, from January 1992 through May 1997, the VDAX showed the lowest volatility levels and an upward moving stock index (a period in a low-volatility bull market). From June 1997 through April 2000, there were high VDAX levels (a bubble period where we find a high-volatility bull market). A spike in the VDAX can be seen in 1998. This was followed by a stock market crash due to the Long-Term Capital Management (LTCM) crisis; during this crash, the VDAX level reached as high as 57%. A crash can also be observed from April 2000 through January 2003 when the bubble burst (a period in a high-volatility bear market). However, from 2004 to August 2007 (a period in a low-volatility bull market), we found positive returns, and the level of the VDAX decreased, showing low volatility levels. From September 2007 through May 2009 (a period of extremely high volatility and an extreme bear market), we found the highest VDAX levels, decreasing stock index returns and some jumps due to the credit crunch and liquidity crises. In this particular period, a spike in the VDAX level was observed in October 2008 when the market crashed and the VDAX level reached a historical peak of 74%.

Table 1 provides the summary statistics for the daily continuously compounded returns (in %) of the DAX 30 index as well as tests for normality and unit roots.

<table>
<thead>
<tr>
<th></th>
<th>DAX 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.000251</td>
</tr>
<tr>
<td>Median</td>
<td>0.000501</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.107975</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.088747</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>0.014682</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.111999</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>8.228179</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>5183.57**</td>
</tr>
<tr>
<td>ADF</td>
<td>-68.75**</td>
</tr>
<tr>
<td>No. Obs</td>
<td>4543</td>
</tr>
</tbody>
</table>

Table 1 reports the descriptive statistics of the returns series of DAX30 index. The Jarque-Bera and the Augmented Dicky Fuller (ADF) (an intercept is included in the test equation) test values are reported. ** and * denote rejection of the null hypothesis at the 1% and 5% significance levels respectively.

The mean return on the DAX 30 index is not statistically different from zero. The tests for
skewness and kurtosis confirm that returns on the DAX 30 index are negatively skewed and are highly leptokurtic with respect to a normal distribution. Likewise, the Jarque-Bera statistics reject normality for the return series. The stationarity in the return series was investigated by applying the augmented Dickey-Fuller (ADF) test for a unit root. The ADF results reject the hypothesis of a unit root in the return series at the 1% significance level.

Figure 2 shows a Q-Q plot of the theoretical quantiles of the normal distribution (vertical axis) against empirical quantiles of the returns on the DAX 30 (horizontal axis). As can be seen from this plot, the data are not normally distributed, indicated by the fact that the empirical quantiles do not lie on a straight line.

![Quantiles of returns on DAX 30](image)

**Figure 2.** Normal Q-Q plots returns on DAX 30.

As can be seen from the plot, there is a significant deviation from the straight line in the tails, particularly in the lower tail, indicating that the distribution of returns on the DAX 30 index is more heavily tailed than the normal distribution. The Q-Q plot thus reinforces the findings of our earlier statistical tests for normality, leading to the use of a heavy-tail distribution rather than a normal distribution in the rest of our analysis.
4 Daily Value-at-Risk Models

4.1 Value-at-Risk

Value at Risk (VaR) is defined as the maximum expected loss in the value of a portfolio. It has a certain probability over a certain holding period (for details on VaR, see, e.g., Duffie and Pan, 1997; Jorion, 2000; Dowd, 2005; Christoffersen, 2009). VaR forecasts are fundamental to financial risk management and risk regulation. However, the importance and recognition of VaR as a risk management tool spring from the Market Risk Amendment (1996) to the Basel Capital Accord of 1998 and also because of the popularity of the RiskMetrics introduced by J.P. Morgan (see Jorion, 2000; Dowd, 2005). Since then VaR became widely accepted by banks and was also imposed by regulators. The aim was to supervise and manage market risk—the market-exposure risk arising due to unfavorable movements in equity prices, interest rates, exchange rates, commodity prices etc. Because VaR has become a standard measure for market risk due to the trading activities and market positions taken by large banks, most financial institutions and trading houses currently use VaR models to assess their daily portfolio losses from significant trading activities. They also backtest VaR models by observing when the portfolio returns exceed the VaR forecasts; that is, the VaR forecasts match their expectations. As a result, accurate VaR forecasts are crucial for market risk management. Given that accurate VaR forecasts rely heavily on the accurate forecasting of the volatility of a portfolio, this is an important parameter for any VaR model.

For instance, the level of the VaR over a one-day holding period is defined as the solution

\[ P(r_t^p < -VaR_t^p \mid \Omega_{t-1}) = \alpha, \]

where \( \alpha \) is 1 minus the VaR confidence level (e.g. 99%), and \( r_t^p \) is the return of a portfolio over a one-day holding period. Having conditional volatility specification \( h_t \) such that

\[ r_t^p = \sqrt{h_t} \epsilon_t, \]

where the residuals are distributed as \( \epsilon_t \sim N(0, h_t) \), then a one-period VaR at time \( t \) is

\[ VaR_t = \Phi^{-1}(\alpha) \sqrt{h_t}, \]

where \( \Phi \) denotes the standardized normal cumulative distribution function. In fact, we need to accommodate heavy tails in VaR estimation; therefore, VaR needs to be estimated using a Student’s \( t \) density function.\(^5\) We assume that \( \epsilon_t \) are distributed as follows:

\(^5\)Previous studies have concluded that distribution functions accounting for fat tails are fundamental for VaR modeling. See, for instance, Huisman et al. (1998), and Alexander and Sheedy (2008) for VaR forecasts obtained through different density functions.
\[ e_i \sqrt{h_t} (v/(v-2))^{1/2} \sim T_v, \]  

where \( T_v \) is the standardized Student’s \( t \)-distribution with \( v \) degrees of freedom.

To estimate the \( \alpha \)% daily VaR using Student’s \( t \) density function: \(^6\)

\[ VaR_{\alpha} = T_{\alpha}^{-1} \left( \frac{(v-2)}{v} \right)^{1/2} \sqrt{h_t}, \]

where the shape \( v \) is to be estimated. Here, we used four different types of volatility forecasts (implied volatility (VDAX), implied volatility (VDAX) with GJR, FHS (GJR), and RiskMetrics) as parameters for the daily VaR models; these are specified and discussed in the following subsections.

### 4.2 Implied Volatility

We considered the new VDAX (MFIV) index, which is a robust volatility measure. The rationale for considering the new VDAX is that the VaR modeling is about extreme events i.e., implying volatility from the extreme outcomes in the options market is of paramount importance to VaR forecasting. These extreme events are embedded in the IV derived from OTM options (see, for instance, Liu et al., 2005; Camara and Heston, 2008; Bates, 2008). Likewise, the new VDAX measure is important because it accounts for volatility skew, which may be induced by the net buying pressure of the OTM puts (see Bollen and Whaley, 2004).

Volatility skew is an obvious phenomenon previously documented by many other researchers and is important to capture in any volatility measure (e.g., Alexander, 2001; Low, 2004; Goncalves and Guidolin, 2006; Badshah, 2008). In addition, information from trading strategies and other shocks is incorporated in the new VDAX index as it accounts for both OTM put and call options. Nevertheless, the new VDAX is implied from both the fear and exuberance embedded in option prices, even though the majority of option traders are very informed and possesses strong skills (see, e.g., Low, 2004 and Chakravarty et al., 2004). Thus, the new VDAX is an excellent choice to be used as a volatility parameter in the daily VaR model to quantify a daily VaR forecast for the DAX 30 stock index portfolio. However, for the daily VaR model, a daily-variance parameter is needed instead of the standard deviation as VaR uses variance as an input, as the VDAX is expressed in annualized standard deviation units. Therefore, transformation is essential; for instance, at time \( t \), we insert

\(^6\) As the returns on the DAX 30 are non-normally distributed; therefore, the VaR forecasts are estimated using Student’s \( t \) density function.
$VDAX_{t-1}$ into the daily VaR model as 

$$h_t = \sigma_{imp,1,t-1}^2,$$  

(8) 

while 

$$\sigma_{imp,1,t-1}^2 = \left(\frac{VDAX_{t-1}}{\sqrt{252}}\right)^2.$$  

(9) 

However, Granger and Poon (2003, 2005) point out that BSIV is biased and is always higher than the actual volatility. They suggest using HV for calibration as 

$$h_t = \alpha + \delta \sigma_{imp,1,t-1}^2,$$  

(10) 

where $\alpha$ and $\delta$ need to be estimated. However, we assert that VDAX is calculated using the MFIV measure, whose IV value should not be subject to model risk, and that this bias should thus not be of great concern; therefore, any of the above two variance measures can be used equally. Furthermore, we have a combined specification for variance using GJR-GARCH (1,1) extended with the lagged IV as 

$$h_t = c_0 + c_1 e_{t-1}^{-2} + c_2 e_{t-1} d_{t-1} + \beta h_{t-1} + \delta \sigma_{imp,1,t-1}^2.$$  

(11) 

This equation has a dummy variable to capture asymmetry. For instance, the dummy variable $d_{t-1}$ is equal to 1 when $e_{t-1} < 0$ and is equal to 0 otherwise. Therefore, estimating an $\alpha%$ daily implied (VDAX) VaR forecast or the combined VaR forecast using implied volatility (VDAX) plus GJR can be done with the following specification: 

$$VaR_{t,\alpha} = T_{\nu}^{-1}(\alpha) \left(\frac{\nu - 2}{\nu}\right)^{\nu / 2} \sqrt{h_t},$$  

(12) 

where $T_{\nu}$ is the standardized Student’s $t$-distribution function, the shape $\nu$ parameter needs to be estimated and $\alpha$ is the quantile (which in our case is 99%, 97.5% or 95%).

### 4.2 Filtered Historical Simulation

As we know, historical simulation (bootstrapping) is a model-free approach that uses the past historical returns of each asset in a portfolio in order to generate future scenarios; therefore, a historical simulation (HS) VaR forecast is based on the empirical distribution of asset returns, which is more realistic and economically appealing for measuring a portfolio’s risk. Pritsker (2006) and Boudoukh et al. (1998) thoroughly discussed the assumptions and limitations of HS, pointing out that HS has many disadvantages. HS-VaR forecasting may be misleading; for instance, the asset returns are not independently, identically distributed (iid), leading to

---

7 We use a similar transformation scale to that of Blair et al. (2001).
the present fat-tailed distribution with time-varying conditional moments and volatility clustering. Barone-Adesi et al. (1998) extend the HS and propose Filtered Historical Simulation (FHS), which overcomes limitations of the HS and is consistent with the empirically observed features in the financial market data. FHS has a major advantage in that we do not need to make any assumptions about the distribution in advance; instead, the past data directly tell us about the distribution. Nonetheless, we fit a GARCH-type model to the data. We based the conditional volatility model on the assumption that the model must produce iid-standardized returns. Finally, we came up with the following GJR-GARCH (1,1) model specification:

\[ r_t = \theta e_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, h_t), \]

\[ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 d_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1}. \]

The above equation has a dummy variable to capture asymmetry. The dummy variable \( d_{t-1} \) is equal to 1 when \( \varepsilon_{t-1} < 0 \) and equals 0 otherwise. However, for forecasting using FHS-VaR, the i.i.d.-standardized residual returns are essential. The residuals are divided by the corresponding daily conditional volatility obtained from the GJR model:

\[ z_t = \frac{\varepsilon_t}{\sqrt{h_t}}. \]

The daily FHS-VaR forecast is estimated by using the VaR specification:

\[ VaR_t = -\sqrt{h_t} \cdot \text{Perc} \left( z_{t-1} \right)^n \cdot \alpha, \]

where \( n \) represents the standardized residuals and \( \alpha \) is the quantile (which in our case is 99%, 97.5% or 95%).

### 4.3 RiskMetrics

RiskMetrics\textsuperscript{TM} is a set of methodologies for measuring market risk (e.g., VaR) developed by J.P. Morgan. The RiskMetrics model assumes that the daily return of a portfolio follows a conditional normal distribution. It assigns greater weight to recent observations and less weight to more distant observations. This declining weighting scheme, through the assertion that volatility tends to change over time in a stable way, might be more reasonable than assuming that it is constant (see e.g., Christoffersen, 2003). The specification of the RiskMetrics model is

---

\textsuperscript{8} FHS combines the characteristics of parametric and non-parametric methods, which account for leptokurtosis, time-varying volatility, asymmetry, and serial-correlations.

\textsuperscript{9} The MA term is inserted to remove serial-correlation in the returns series.
where $\lambda$ is weight, which is usually fixed at 0.94 by the RiskMetrics group. The RiskMetrics model’s forecast for one-day-ahead volatility at time $t$ is therefore a weighted average of volatility at time $t-1$ times the squared return at $t-1$. Moreover, this model has several advantages, some of which are worth mentioning here. First, it accounts for time variations in the variance in a manner consistent with observed returns. Second, recent observations are given more importance than the older observations. Third, fewer observations are needed in order to forecast a one-day-ahead variance. Finally, there is an agreement on the parameter value $\lambda = 0.94$ across assets for one-day-ahead variance forecasting (for more details, see, e.g., Christoffersen, 2003). However, we improve the RiskMetrics VaR forecast by accommodating a heavy-tailed Student’s $t$-distribution function instead of a normal distribution function. Thus, the daily $\alpha\%$ RiskMetrics VaR forecast is estimated as

$$VaR_{t,\alpha} = T^{-1}_\nu(\alpha) \left( \frac{v - 2}{v} \right)^{1/2} \sqrt{h_t},$$

where $T_\nu$ is the standardized Student’s $t$-distribution function, the shape $\nu$ parameter needs to be estimated, and $\alpha$ is the quantile (which in our case is 99%, 97.5% or 95%).

5 Backtesting Daily VaR Models

The current regulatory framework requires financial institutions with massive trading activities to have enough capital, called a market capital requirement (MRC), in order to cover excessive portfolio losses. This MRC requirement is determined in terms of the portfolio VaR measure. The regulatory frameworks also necessitate that financial institutions should use their own internal VaR models to compute and provide their 99% VaR. As a result, the MRC requirements are directly linked to both the portfolio risk level and the internal VaR model’s performance on backtests. However, based on VaR model forecasts, usually the regulatory body increases (decreases) MRC requirements (for more details on MRC, see Campbell, 2006). Thus, the accuracy of internal VaR models is of paramount importance to both regulators and financial institutions.\(^{10}\) When examining the accuracy of a VaR model, the significance of the backtesting technique used increases. In this respect, Kupiec (1995) was the first to propose a test of unconditional coverage; later, Christoffersen

\(^{10}\)For instance, if a VaR model indicates less risk, then the regulatory body could reduce MRC requirements, and the financial institution could then use that capital for other financial activities. Thus, accurate VaR forecasting is important for a financial institution.
(1998) extended the Kupiec test to the test of conditional coverage.

Let’s assume we want to backtest a VaR model; to do that, a hit sequence needs to be defined. For instance, if ex-ante VaR forecasts and ex-post returns are observed in a time-series manner, then a hit sequence of VaR violations (or exceptions) are represented with the following indicative function:

\[
I_t(\alpha) = \begin{cases} 
1, & \text{if } r_t^p < -VaR_t^p(\alpha) \\
0, & \text{if } r_t^p > -VaR_t^p(\alpha)
\end{cases}
\]

(19)

If a portfolio’s loss on day \( t \) is greater than the forecasted VaR for day \( t \), then the hit sequence reports 1, implying a VaR violation; otherwise, the hit sequence reports 0, implying no VaR violation. In this way, the hit sequence \( \{I_t\}_{t=1}^T \) across \( T \) days is constructed, indicating the violation rate of the VaR model. In fact, Christoffersen (1998) suggests that the VaR model is adequate when its ‘hit sequence’ satisfies both unconditional coverage and independence properties. \(^{11}\) The unconditional coverage property: The probability that an ex-post loss exceeds VaR forecasts must be exactly equal to the coverage rate. The independence coverage property: The VaR violations observed at two different periods must be independently distributed over time. The observed VaR violations do not carry information to forecast current and future VaR violations, and this property also holds for any other variable in the information set, for instance, past returns, past VaR levels, etc. Therefore, the null hypothesis of the unconditional coverage property that \( E[I_t] = P \) can formally be tested using the log-likelihood ratio test is

\[
LR_{UC} = 2\ln \left( \left(1 - \frac{X}{T}\right)^{T-X} \left(\frac{X}{T}\right)^{X} \right) - 2\ln \left( (1-P)^{T-X} \right) \sim x^2_1,
\]

(20)

which is asymptotically \( x^2 \)-distributed with one degree of freedom. \( P \) is the target coverage rate (i.e., 1% for a 99% VaR), \( T \) is the total number of observations, and \( X \) is the number of violations. Unfortunately, the unconditional coverage test does not account for the clustering of violations. \(^{12}\) To overcome a clustering problem, the independence coverage property must be satisfied, thereby leading to the correct conditional coverage test proposed by Christoffersen (1998): joint tests for both independence and unconditional coverage tests. \(^{13}\)

\(^{11}\) Christoffersen (2009) thoroughly reviews backtesting techniques.

\(^{12}\) The unconditional coverage test has a major problem as it does not account for the clustering of ones or zeros in a time-dependent fashion. For instance, if a value of 1% gave exactly 1% violations, but all of these violations occurred during one-month period, then the probability of bankruptcy of an institution would be much higher than if the VaR violations are scattered over a longer backtesting period, such as a one or two-year period.

\(^{13}\) For some technical details, see Christoffersen (1998) and Campbell (2005).
The likelihood ratio tests for these tests are

\[
LR_{\text{IND}} = 2 \ln \left( \left( I - \pi_0 \right)^{T_{ij}^{0} \pi_0^{T_{ij}^{0}}} \left( I - \pi_1 \right)^{T_{ij}^{1} \pi_1^{T_{ij}^{1}}} \right) - 2 \ln \left( \frac{I - X}{T} \right)^{T_{ij} \pi_0^{T_{ij}}} \left( \frac{X}{T} \right)^{T_{ij} \pi_1^{T_{ij}}} \right) \sim \chi^2, \quad (21)
\]

\[
LR_{\text{CC}} = LR_{\text{LUC}} + LR_{\text{IND}}, \quad (22)
\]

where \( T_{ij} \) is the number of observations, with value \( i \) followed by \( j \), for \( i, j = 0, 1 \); \( \pi_{ij} = T_{ij} / \sum_j T_{ij} \) are the corresponding probabilities; \( ij = 1 \) indicates that a violation has occurred; and \( ij = 0 \) indicates the opposite. The \( LR_{\text{CC}} \) statistics are asymptotically \( \chi^2 \) distributed with two degrees of freedom. The null hypothesis in the independence test \( LR_{\text{IND}} \) states that the probability of a violation on a given day does not depend on the previous day’s violation.

An alternative method for evaluating VaR forecasts could be based on the use of loss functions that are more aligned to the regulatory requirements. Lopez (1999) was the first to propose a loss function for evaluating VaR forecasts, here called Lopez-I. However, this loss function has a major drawback in that it ignores the magnitude of tail losses. To remedy this drawback, he himself proposed a second loss function called Lopez-II. In practice, Lopez-II has a size-adjusted loss function that makes it hard to interpret in monetary terms because it employs squared monetary returns.14 Later, Dowd (2005) proposed a size-based loss function that avoids terms in the denominator and the squared term in Lopez-II; therefore, we prefer Dowd’s size-based function over the others. It takes the following form:

\[
C_i = \begin{cases} 
L_i & \text{if } L_i > VaR_i, \\
0 & \text{if } L_i \leq VaR_i,
\end{cases} \quad (23)
\]

where \( L_i \) is the loss incurred over period \( t \) and \( VaR_i \) is the forecasted VaR estimate for that period. We then calculated the expected tail loss \( ET \) by using \( C_i \), and \( ET \) was used as a benchmark. Finally, the quadratic score function takes the following form:

\[
QS = \frac{2}{n} \sum_{i=1}^{n} \left( C_i - ET_i \right)^2, \quad (24)
\]

This function has advantages over others because it punishes deviations of the tail losses from the expected value and because of its quadratic form; it gives very high tail losses much greater weight than normal tail losses.

14 See Dowd (2005) for detailed discussions on different loss functions.
6 Results

6.1 Conditional Variances

Figure 3 provides a comparison of the conditional variances for implied volatility (VDAX), GJR plus implied volatility (VDAX), GJR-GARCH (1, 1), and RiskMetrics specifications and the squared returns that they forecast. The comparison period for the five time series is from January 1, 1992 through May 29, 2009. It can be observed that the squared returns are very noisy and that the four variances move closely together.

![Figure 3](image)

**Figure 3.** Implied, conditional variances and squared returns from January 01, 1992 to May 29, 2009.
6.2 Backtesting Results

Table 2 provides backtesting results of VaR models for the DAX 30 stock index portfolio from January 1, 1992 to May 29, 2009. Implied volatility (VDAX), implied volatility–GJR, FHS (GJR) and RiskMetrics VaR models were statistically tested using tests of unconditional coverage, independence, and conditional coverage and the quadratic score test (using the expected tail loss) at different confidence levels, for example, 99%, 97.5% and 95%. There are three panels in Table 2. In rows 1, 2, and 3 of each panel, the results of the unconditional coverage, independence and conditional coverage tests are provided, respectively. In rows 4, 5, 6, 7, and 8 of each panel, the number of exceptions, the %VaR exceptions, the number of successive exceptions, the % expected tail-loss, and the quadratic-score are presented, respectively.

First, we discuss the backtesting results for our four specified VaR models at the 99% confidence level, which are provided in panel A of Table 2. As the null hypothesis of the unconditional coverage test states that the observed frequency of exceptions should equal the frequency of expected exceptions, therefore testing unconditional coverage is the first step to take when considering any VaR model. As can be seen, the chi-square values for the null hypothesis of the unconditional coverage tests at 10% significance cannot be rejected for any of the four VaR models. However, the unconditional test only considers the frequency of exceptions, and it ignores the time-series independence of those exceptions, i.e., the clustering of exceptions. As a result, the independence test is the most important for any VaR model.

Conducting the independence test is the second step to take when considering any VaR model. The chi-square values at the 10% significance level for the independence tests confirm that the null hypothesis of the first-order independence of exceptions is not rejected for the three specified VaR models, i.e., implied volatility (VDAX)-GJR, FHS (GJR), and RiskMetrics; however, this test could not be conducted for the implied volatility (VDAX) model because there are no successive exceptions implying that the model is correct.

---

15 All VaR models except FHS are estimated in conjunction with the Student’s t density function.
16 The daily 99% VaR forecasts for each of the four VaR models are compared with the corresponding daily returns in a time-series fashion in Figure 4.
17 A chi-square critical value of 2.70 (for one degree of freedom) for a 10% significance level is used to test the null hypothesis that the frequency of observed exceptions is consistent with the frequency of expected exceptions.
### Table 2
Backtesting results of VaR models for the DAX 30 from January 1, 1992 to May 29, 2009.

Panel A: 99% VaR-t forecasts

<table>
<thead>
<tr>
<th>Model</th>
<th>Implied (VDAX)</th>
<th>Implied (VDAX)-GJR</th>
<th>FHS (GJR)</th>
<th>RiskMetrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR_{u}</td>
<td>1.299692</td>
<td>0.663502</td>
<td>0.919916</td>
<td>2.317618</td>
</tr>
<tr>
<td>LR_{l}</td>
<td>-</td>
<td>-6.763572</td>
<td>2.115715</td>
<td>1.698050</td>
</tr>
<tr>
<td>LR_{cc}</td>
<td>-</td>
<td>-6.100069</td>
<td>3.035631</td>
<td>4.015668</td>
</tr>
<tr>
<td>No. of Exceptions</td>
<td>38</td>
<td>51</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td>%VaR Exceptions</td>
<td>0.843507</td>
<td>1.122606</td>
<td>1.144618</td>
<td>1.232666</td>
</tr>
<tr>
<td>No. Succ. Exceptions</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Expected Tail Loss (%)</td>
<td>-4.323369</td>
<td>-3.5817592</td>
<td>-3.71932664</td>
<td>-3.48212508</td>
</tr>
<tr>
<td>Quadratic Score</td>
<td>0.000371</td>
<td>0.000594</td>
<td>0.000635</td>
<td>0.000740</td>
</tr>
</tbody>
</table>

Panel B: 97.5% VaR-t forecasts

<table>
<thead>
<tr>
<th>Model</th>
<th>Implied(VDAX)</th>
<th>Implied(VDAX)-GJR</th>
<th>FHS(GJR)</th>
<th>RiskMetrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR_{u}</td>
<td>0.285262</td>
<td>0.018263</td>
<td>0.781421</td>
<td>3.912869</td>
</tr>
<tr>
<td>LR_{l}</td>
<td>-6.094480</td>
<td>-6.926530</td>
<td>-7.03222</td>
<td>-6.738680</td>
</tr>
<tr>
<td>LR_{cc}</td>
<td>-5.809210</td>
<td>-6.908270</td>
<td>-6.2508</td>
<td>-2.825810</td>
</tr>
<tr>
<td>No. of Exceptions</td>
<td>108</td>
<td>115</td>
<td>123</td>
<td>135</td>
</tr>
<tr>
<td>%VaR Exceptions</td>
<td>2.435175</td>
<td>2.531367</td>
<td>2.707462</td>
<td>2.971605</td>
</tr>
<tr>
<td>No. Succ. Exceptions</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Expected Tail Loss (%)</td>
<td>-3.728668</td>
<td>-3.221964</td>
<td>-3.099948</td>
<td>-3.176138</td>
</tr>
<tr>
<td>Quadratic Score</td>
<td>0.000897</td>
<td>0.001082</td>
<td>0.001288</td>
<td>0.001403</td>
</tr>
</tbody>
</table>

Panel C: 95% VaR-t forecasts

<table>
<thead>
<tr>
<th>Model</th>
<th>Implied(VDAX)</th>
<th>Implied(VDAX)-GJR</th>
<th>FHS(GJR)</th>
<th>RiskMetrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR_{u}</td>
<td>0.123797</td>
<td>1.605154</td>
<td>1.786875</td>
<td>8.434941</td>
</tr>
<tr>
<td>LR_{l}</td>
<td>-4.187958</td>
<td>-1.675680</td>
<td>1.828643</td>
<td>0.047557</td>
</tr>
<tr>
<td>LR_{cc}</td>
<td>-4.064160</td>
<td>-0.070530</td>
<td>3.615519</td>
<td>8.482498</td>
</tr>
<tr>
<td>No. of Exceptions</td>
<td>222</td>
<td>246</td>
<td>247</td>
<td>271</td>
</tr>
<tr>
<td>%VaR Exceptions</td>
<td>5.137699</td>
<td>5.414924</td>
<td>5.436936</td>
<td>5.965221</td>
</tr>
<tr>
<td>No. Succ. Exceptions</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>Expected Tail Loss (%)</td>
<td>-3.295459</td>
<td>-2.841407</td>
<td>-2.783163</td>
<td>-2.751958</td>
</tr>
<tr>
<td>Quadratic Score</td>
<td>0.001844</td>
<td>0.002288</td>
<td>0.002334</td>
<td>0.002613</td>
</tr>
</tbody>
</table>

Boldface indicates that we cannot reject the null hypothesis at 10% significance level.
Finally, the conditional coverage test that needs to be tested in order to endorse any VaR model is the joint null hypothesis that the VaR model has an adequate frequency of independence exceptions. This test also provides important information regarding the likelihood of the successive exceptions and the average number of days between exceptions; this is the third and final step for a VaR model to be considered for use. As can be observed, the chi-square values at the 10% significance level for the joint null hypothesis tests verify that the three VaR models have an adequate frequency of independence exceptions; as a result, the joint null hypothesis could not be rejected. Similarly, for the implied volatility (VDAX) model, the conditional coverage test could not be conducted because there are no successive exceptions implying that this model is correct under the conditional coverage test.18

For comparison, the backtesting results for the VaR models at 97.5% and 95% confidence levels are provided in panels B and C of Table 2.19 As can be seen, the chi-square values for the null hypothesis of the unconditional coverage tests at the 10% significance level also could not be rejected for three of the VaR models. However, in the case of the RiskMetrics model, we found a higher unconditional coverage than expected, leading to a rejection of the null hypothesis of correct unconditional coverage at the 10% significance level. In comparison, the chi-square values at 10% significance for the independence tests show that the null hypothesis of the independence of the exceptions was not rejected for our VaR models. As can be observed, for the conditional coverage test, the chi-square values at the 10% significance level for the joint null hypothesis tests verify that the four VaR models have an adequate frequency of independence exceptions; therefore, the joint null hypothesis could not be rejected. However, the null hypothesis for the conditional coverage test was rejected at the 10% significance level for the RiskMetrics model (at the 95% confidence level).

In order to select a model, we can also use estimated quadratic scores for each VaR model at different confidence levels such as 99%, 97.5% and 95%. However, the quadratic scores are estimated using a loss function proposed by Dowd (2005), which states that a lower minimum score level indicates a better VaR model. The quadratic scores are presented in the last rows of each panel of Table 2. If we compare the quadratic scores for each of the VaR models at the 99% confidence level, the rank of the VaR models would be, from better to worse, implied volatility (VDAX), implied volatility (VDAX)-GJR, FHS (GJR), and

---

18 Chi-square critical value of 4.60 (for two degree of freedom) for 10% significance level is used to test the joint null hypothesis of the conditional coverage test.

19 The daily 97.5% and 95% VaR forecasts for each of the four specified VaR models are compared with corresponding daily returns in a time-series fashion in Figure 5, and 6, respectively.
RiskMetrics. Alternatively, in a more traditional way, one could also select a VaR model by evaluating it based on the number of exceptions and the number of successive exceptions it has previously made. By these criteria, the VaR models’ rank is similar to the rank we determined from the quadratic score estimates. In addition, ranks for the VaR models at lower confidence levels (97.5% and 95%) support the quadratic score estimates of the VaR models at the 99% confidence level.

We conclude from the results presented in Table 2 that the information content of IV is superior to that of HV, which is consistent with the earlier research on the information content (e.g., Day and Lewis, 1992; Christensen and Prabhala, 1998; Fleming, 1998; Dumas, Fleming and Whaley, 1998; Blair et al., 2001; Ederington and Guan, 2002; Poon and Granger, 2003; Mayhew and Stivers, 2003; Martens and Zein, 2004; and Giot, 2005). The unconditional coverage, independence, and conditional coverage tests and the quadratic scores suggest that the VDAX and a combination of the VDAX with GJR-GARCH (1,1) VaR models always outperform filtered historical simulation (FHS) and RiskMetrics models during our sample period from January 1, 1992 through May 29, 2009.
7 Conclusion

This paper examined the information content of the newly adopted VDAX (MFIV) index for the daily VaR forecasts. The information content of the new VDAX was incorporated into daily VaR forecasts and compared with the VaR forecasts from the FHS (GJR) and the RiskMetrics models at various confidence levels (i.e., 99%, 97.5% and 95%), using unconditional coverage, independence, and conditional coverage tests for backtesting each of the VaR models. Furthermore, a quadratic score was estimated for each VaR model for the period from January 1, 1992 through May 29, 2009. The backtesting results showed that the new VDAX index contains significant information about actual volatility in VaR models. The null hypotheses of independence and conditional coverage backtests were never rejected for implied volatility, implied volatility plus GJR, and FHS (GJR) VaR models. The number of VaR exceptions was not significantly different from the set coverage rates. However, the null hypotheses of the RiskMetrics model were rejected for lower confidence levels. It was also found that implied volatility and implied volatility-GJR VaR models presented the fewest VaR exceptions and clusters of exceptions, in contrast to the FHS (GJR) and RiskMetrics models. On the other hand, the quadratic score for each model suggests the following ranking of VaR models: implied volatility (VDAX), combined (implied volatility plus GJR), FHS, and RiskMetrics. Our findings have implications for traders who hold long positions, risk managers (internal), and regulators (external).
References

Ederington, L., and Guan, W., 2002. Is implied volatility an informationally efficient and


Figure 4. Daily 99% VaR forecasts versus returns on DAX 30 index from January 01, 1992 to May 29, 2009.
Figure 5. Daily 97.5% VaR forecasts versus returns on DAX 30 index from January 01, 1992 to May 29, 2009.
Figure 6. Daily 95% VaR forecasts versus returns on DAX 30 index from January 01, 1992 to May 29, 2009.