



Swansea University  
Prifysgol Abertawe



## Cronfa - Swansea University Open Access Repository

---

This is an author produced version of a paper published in:  
*Journal of Futures Markets*

Cronfa URL for this paper:  
<http://cronfa.swan.ac.uk/Record/cronfa34904>

---

### **Paper:**

Kambouroudis, D., McMillan, D. & Tsakou, K. (2016). Forecasting Stock Return Volatility: A Comparison of GARCH, Implied Volatility, and Realized Volatility Models. *Journal of Futures Markets*, 36(12), 1127-1163.  
<http://dx.doi.org/10.1002/fut.21783>

---

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

# Forecasting stock return volatility: a comparison of GARCH, implied volatility and realized volatility models

Dimos Kambouroudis, David McMillan, Katerina Tsakou

October 28, 2015

## Abstract

This paper investigates the information content of implied volatility forecasts in the context of forecasting stock index return volatility by studying a number of US and European indices. Using a number of different autoregressive models for forecasting implied volatility, we examine whether implied volatility forecasts contain any additional information useful to predict future volatility beyond that embedded in GARCH models and realized volatility. The results show that implied volatility follows a predictable pattern and confirms previous literature that there is a contemporaneous relationship between implied volatility and index returns. When the predictive power of the implied volatility for future volatility is assessed, it is found that, overall, implied volatility contains additional information compared to GARCH and realized volatility, although individually performs worse. Nevertheless, a model that combines the information contained in an asymmetric GARCH model with the information from option markets and realized volatility through (asymmetric) ARMA models is the most appropriate for predicting future volatility. This evidence is also supported in the context of value-at-risk.

## 1 Introduction

Modelling and forecasting volatility is an important task in financial markets. Over the past few decades there is an extensive research agenda that has analyzed the importance of volatility in investment, option pricing and risk management. Thus, accurate estimation and forecasting of asset returns volatility is crucial for assessing investment risk.

Broadly speaking, there are two approaches that the majority of researchers adopt to generate volatility forecasts. The first method is to extract information about the variance of future returns from historical data using simple models, GARCH-type models, stochastic volatility models or the

realized volatility. The second method is to extract market expectations about future volatility from observed option prices, using implied volatility (henceforth IV) indices. The focus of this paper lies on the GARCH-type models, implied volatility and realized volatility.

Early empirical research (Latane and Rendleman, 1976; Chiras and Manaster, 1978) support the idea of using implied volatility as a predictor for future volatility. In subsequent research, and among others, Day and Lewis (1992) compare the information content of IV for the S&P100 index options to GARCH type conditional volatility and find that IV models contain predictive information about future volatility beyond that contained in GARCH. A similar conclusion has been reached by Lamoureux and Lastrapes (1993) who study several individual stocks. But the findings in these studies are subject to potential measurement errors due to overlapping samples and maturity mismatch problems. Overcoming these problems, more recent papers favour the conclusion that IV is informationally efficient in forecasting future volatility. For example, Christensen and Prabhala (1998) utilize non-overlapping samples to study S&P100 index options and document that IV outperforms historical volatility. Using daily index returns and/or intraday returns Blair et al. (2001) for the S&P100 index and the VIX find that VIX provides more accurate forecasts than GARCH-type models, in particular, as the forecast horizon increases. A combination of VIX and GJR-GARCH forecasts are more informative than VIX and GJR-GARCH alone when forecasting one-day ahead. They also find that high-frequency returns contain useful information about future volatility and provide more accurate measure of realized volatility than daily returns.

The accuracy of volatility forecasting has always been the subject of extensive research. Since the Chicago Board Exchange (CBOE) launched VIX, literature that compares options based forecasts with those from time series models is voluminous. Nonetheless, no conclusion has been reached yet and hence, there is still an ongoing debate between GARCH-type models and IV index models in finding the preferred model in estimating and forecasting future volatility. Furthermore, the availability of intraday data and the construction of realized volatility (RV) by (Andersen and Bollerslev, 1998; Andersen et al., 1999) has improved the method of measuring and forecasting volatility. (Andersen and Bollerslev, 1998; Andersen et al., 1999) show that RV, defined as the sum of squared intraday returns, provides a more accurate measure of 'true volatility' than daily squared returns.

For the German economy, Claessen and Mittnik (2002) find that, although the null hypothesis that the German IV index (VDAX) is an unbiased estimate for realized volatility is rejected, GARCH volatility does not contain useful information beyond the volatility expectations already reflected in option prices. Giot (2005a) and Corrado and Miller (2005) conclude that the volatility forecast based on the VIX and VXN indices, i.e. the IV index based on NASDAQ100 index, have the highest information content both for volatility forecasting and for market risk assessment

framework. However, Giot (2005a) concludes that combining GARCH with implied volatility often improves on the results from either one alone. Carr and Wu (2006) for the S&P500 stock index, Yu et al. (2010) using stock index options traded over-the-counter and on exchanges in Hong Kong and Japan and Yang and Liu (2012) for the Taiwan stock index reach similar conclusions. Frijns et al. (2010), for the Australian index, find that at short horizons combining GJR-GARCH and IV improve future volatility forecast, but overall IV outperforms the RiskMetrics and GJR-GARCH. In a similar vein, Cheng and Fung (2012) show that while IV is more informative than GARCH, the GARCH forecast improves the predictive ability of IV for the Hong Kong market. On the other hand, and among others, the results of Becker et al. (2007) contradict the previous studies, because they show that VIX is not an efficient volatility predictor and does not provide any additional information relevant to future volatility.

Another approach to examining the nature of GARCH and IV forecasts is through forecast combination, for which Clemen (1989) reviews more than 200 papers. Subsequently, a number of studies comparing the performance of different combining methods and conclude that combination approaches usually outperform individual forecasts, for example Diebold and Lopez (1996), Hendry and Clements (2004), Timmermann (2006), Kim and Swanson (2014) and Cheng and Hansen (2015). One of the earlier studies measuring implied volatility using the S&P500 futures and options index is by Ederington and Guan (2002), where they adopt an averaging approach on forecast combinations taking into account averages proposed by both the finance literature as well as by practitioners. More recently, Tsai et al. (2015), investigate the informational content of the VIX options market and how this information is incorporated into the underlying asset price, in this case the VIX index.

Realized volatility is not only considered as a proxy for latent volatility but also as a variable that can be modelled directly (Andersen et al., 2003). Among others, (Martens and Zein, 2004) examine the forecasting value of GARCH, RV and IV show that IV subsumes almost all the information contained in GARCH. Nonetheless, when IV competes with RV through a long memory model, they both contain useful information for forecasting future volatility. Moreover, (Koopman et al., 2005) find that forecasting RV by an ARFIMA model produce more accurate forecasts than GARCH for the S&P100.

There is also another distinct and more recent strand of literature that investigates whether the dynamics of implied volatility *per se* can be forecast. Ahoniemi (2006) uses linear and probit models to model the VIX index. The author finds that an ARIMA(1,1,1) model enhanced with exogenous regressors performs well. The use of GARCH terms in the ARIMA(1,1,1) model are statistically significant, but do not improve the forecast accuracy of the model. Konstantinidi et al. (2008) examine five alternative model specifications to form both point and interval forecasts using a number of US and European IV indices. They find that the ARIMA(1,1,1) and ARFIMA(1,d,1)

specifications provide the best point forecast for the US indices. In a similar spirit, Dunis et al. (2013) investigate the forecasting ability of intraday IV on an underlying EUR-USD exchange rate for a number of maturities by combining a variety of forecasting models. They find that the GJR-GARCH model and the principal component model perform better for one-month and three-months maturity, while ARFIMA and VAR models outperform for longer periods. Finally, Fernandes et al. (2014) perform a thorough statistical examination of the time series properties of the VIX. The out-of-sample analysis shows that ARMA models perform very well in the short run and very poorly in the long run, while the semi-parametric heterogeneous autoregressive (HAR) process perform relatively well across all forecasting horizons.

This paper participates in the ongoing debate about the forecasting power of IV for predicting realized volatility making several contributions. First, motivated by the recent research on the forecasting ability of IV itself, this paper investigates whether the IV forecast generated by ARMA-type models are good predictors of stock index volatility providing evidence from the S&P500, DJIA and NASDAQ100 indices. Second, and in order to examine whether the results may differ across markets, we repeat our analysis using six European indices. The predictability of the IV is assessed by comparing the ARMA type models against a random walk benchmark. Third, the forecasting value of GARCH, IV and RV is assessed. In total, ten GARCH models are considered, GARCH, GJR-GARCH, EGARCH, APGARCH and ACGARCH and their augmented specifications adding the lagged value of the implied volatility. For forecasting IV indices and RV ARMA, ARIMA and ARFIMA models and their unrestricted specifications for capturing the asymmetric relationship between stock index returns and volatility are considered. The results show that all model classes contain important information for predicting future volatility. Asymmetry is important for both GARCH and IV models both in-sample and out-of-sample. Overall, for the US data, a model which combines the information contained in ACGARCH forecast with the information from option markets and information of RV through ARFIMAX models is the most appropriate for predicting future volatility. As for the EU data, a mixture of an asymmetric GARCH, IV and RV forecast through an ARMAX model performs best.

The remainder of the paper is organized as follows. The next section introduces the data and describes the models and the methodology employed. Section 3 presents the empirical results and Section 4 the results obtained using European data. Section 5 considers an application of the volatility forecasts to risk management practice (VaR), and, finally, summarizes and concludes.

## 2 Data and Empirical Methodology

### 2.1 Data

The dataset primarily consists of the daily closing price data for the S&P Composite 500 (S&P500), Dow Jones Industrial Average (DJIA) and Nasdaq100 indices and their implied volatility indices, VIX, VXD and VXN, respectively. Since the various implied volatility indices have been listed on different dates, we consider the period from February 2, 2001 to February 28, 2013 in order to study the indices over the same time period. The in-sample period is from February 2, 2001 to February 23, 2010 and the remaining period will be used for the out-of-sample evaluation. As volatility is latent, the realized variance is used as proxy for the true volatility. The data of the realized variance are taken from Oxford-Man Institute's Realized Library version 0.2, Heber et al. (2009).

Table 1 presents the descriptive statistics of the stock market returns and the IV indices (Panel A and B, respectively). Panel A shows that the mean and the median are consistently close to zero. As far as the values of skewness and kurtosis are concerned, for a normal distribution, they should be zero and three, respectively. The negative skewness of all series indicates asymmetric distributions skewed to the left, while the kurtosis statistics show the leptokurtic characteristic of all returns distributions. The evidence of non-normality is further supported by the Jarque-Bera test statistic which rejects the null hypothesis of normal distribution at the 1% level. Similarly, Panel B shows the summary statistics of the IV indices along with Augmented Dickey-Fuller (ADF) test for unit roots. The p-values of the ADF tests show that implied volatility indices are stationary at conventional levels. As the IV indices are quoted in annualized percentages and, following Blair et al. (2001), the daily implied index volatility is equal to  $\frac{IV}{100*\sqrt{252}}$ .

### 2.2 Volatility forecasting models

#### GARCH models

The return process is given by

$$r_t = \mu + \varepsilon_t \tag{1}$$

where  $\mu$  is the constant mean and  $\varepsilon_t = h_t z_t$  is the innovation term with  $z_t \sim N(0, 1)$ .

The generalized ARCH (GARCH) model developed by Engle (1982) and Bollerslev (1986) involves a joint estimation of the mean equation (1) and the conditional variance equation as follows

$$h_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2 \quad (2)$$

The empirical success of the GARCH model triggered the development of models that accommodate important stylized facts, such as the asymmetric effect and the long-memory. The GJR-GARCH model has been proposed by Glosten et al. (1993) and is specified as:

$$h_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{t-1} \quad (3)$$

where the leverage effect is captured by the dummy variable  $I_{t-1}$ , such that  $I_{t-1} = 1$  if  $\varepsilon_{t-1} < 0$  and  $I_{t-1} = 0$  if  $\varepsilon_{t-1} > 0$  and the significance of  $\gamma$ .

The exponential GARCH (EGARCH) model has been proposed by Nelson (1991) in order to capture the leverage effect. Nelson (1991) used the EGARCH model to model daily returns of the CRSP value-weighted stock market index in the period 1962-1987. Nelson confirmed that returns are significantly negatively correlated with volatility. The EGARCH model is given by

$$\ln(h_t^2) = a_0 + a_1 \frac{|\varepsilon_{t-1}|}{h_{t-1}} + \gamma \frac{\varepsilon_{t-1}}{h_{t-1}} + \beta_1 \ln(h_{t-1}^2) \quad (4)$$

where the coefficient  $\gamma$  captures the presence of the leverage effects if  $\gamma < 0$ .

In order to investigate the log-run and short-run movement of volatility the component GARCH (CGARCH) model has been developed by Engle and Lee (1993). While the GARCH model and its asymmetric extensions show mean reversion to the unconditional variance, which is constant for all time, the CGARCH model allows mean reversion to a time-varying long-run volatility level,  $q_t$ . The specification of the CGARCH model is given by:

$$h_t^2 = q_t + a_1 (\varepsilon_{t-1}^2 - q_{t-1}) + \beta_1 (h_{t-1}^2 - q_{t-1}) \quad (5)$$

$q_t = a_0 + \rho q_{t-1} + \phi (\varepsilon_{t-1}^2 - h_{t-1}^2)$  is the time-varying long-run volatility provided  $\rho > (a_1 + \beta_1)$ . The forecast error  $(\varepsilon_{t-1}^2 - q_{t-1})$  drives the time-varying process of  $q_t$  and the difference between the conditional variance and its trend,  $(h_t^2 - q_t)$ , is the transitory or short-run component of the conditional variance. Stationarity is accomplished provided  $(a_1 + \beta_1)(1 - \rho) + \rho < 1$ , which in turn requires  $\rho < 1$  and  $a_1 + \beta_1 < 1$ .

An extension is provided by the Asymmetric Component GARCH, ACGARCH, which incorporates the asymmetric dynamics of the GJR model

$$h_t^2 = q_t + (\alpha_1 - \gamma I_{t-1}) (\varepsilon_{t-1}^2 - q_{t-1}) + \beta_1 (h_{t-1}^2 - q_{t-1}) \quad (6)$$

The asymmetric effect is captured by the dummy variable  $I_{t-1}$ , such that  $I_{t-1} = 1$  if  $\varepsilon_{t-1} < 0$

and  $I_{t-1} = 0$  if  $\varepsilon_{t-1} > 0$  and the significance of  $\gamma$ .

These models are also augmented with the lagged of IV in order to assess whether IV contains any incremental information beyond that contained in the GARCH specification and useful for estimating and forecasting the conditional variance.

### ARMA models

In order to investigate whether implied volatility *per se* can be forecast and whether the IV index model forecast will be more accurate than the GARCH type models different ARMA models are going to be used. The choice of the ARMA form is motivated by the previous work of Ahoniemi (2006) and Konstantinidi et al. (2008) who both report some empirical success using such models.

For each IV index an ARMA(1,1) is employed of the form

$$IV_t = c_0 + \phi_1 IV_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (7)$$

A generalization of the ARMA models is the autoregressive integrated moving average (ARIMA) model. It is usually denoted as ARIMA(p,d,q) and is employed to capture the possible presence of short memory features in the dynamics of implied volatility. The ARIMA(1,1,1) specification is given by

$$\Delta IV_t = c_0 + \phi_1 \Delta IV_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (8)$$

In order to account for the possible presence of long memory, and, following Konstantinidi et al. (2008) and Dunis et al. (2013), we also apply a fractionally integrated ARMA model defined by

$$\phi(L)(1-L)^d IV_t = c_0 + \theta(L)\varepsilon_t \quad (9)$$

where  $d$  dictates the order of fractional integration and takes non-integer values. If  $|d| < 0.5$ , the ARFIMA is both stationary and invertible. In particular, if  $d \in (0, 0.5)$ , the process is said to exhibit long memory, while if  $d \in (-0.5, 0)$ , the process exhibits antipersistence. The ARFIMA(1,d,1) model is employed based on the BIC criterion and estimated by maximum likelihood.

Similar to previous studies, such as Giot (2005b), we also examine the possible contemporaneous asymmetric relationship between the index returns and the IV. We add  $r_t^+$  and  $r_t^-$ , the positive and negative underlying index returns, as exogenous variables in the ARMA specifications.

We assess the predictability of IV itself by comparing the above mentioned forecasting model against the random walk benchmark.

$$IV_t = IV_{t-1} + \varepsilon_t \quad (10)$$



Finally, we also use the same ARMA-type models in order to forecast the realized volatility series.

### 2.3 Forecast evaluation

The forecasts are obtained recursively by increasing the sample length by one observation. In other words, the initial estimation date is fixed and, once a forecast is obtained, the end sample date is expanded by one observation and the forecast models re-estimated. We choose a recursive over the alternative rolling approach as the former provides more accurate estimates as the sample size increases. Such increased reliability of the recursive estimates as the sample expands in turns leads to more reliable forecasts.

For examining the forecasting ability of IV itself, the Conditional Predictive Ability (CPA) test of Giacomini and White (2006) is employed. This test evaluates the forecasting performance of two competing models, accounting for parameter uncertainty. In short, let  $L(y_t; \hat{y}_t)$  denote the forecast loss where  $y_t$  is the 'true' value and  $\hat{y}_t$  is the predicted value. The difference in loss of model  $i$  relative to a benchmark model  $o$  is defined as

$$d_{i,t} = L(y_t; \hat{y}_{o,t}) - L(y_t; \hat{y}_{i,t}) \quad (11)$$

The issue is whether the two models have equal predictive ability. That is, the null hypothesis tested is  $H_0 : E(d_{i,t+\tau} | h_t) = 0$ , where  $h_t$  is some information set. The CPA test statistic is then computed as a Wald statistic

$$CPA_t = T(T^{-1} \sum_{t=1}^{T-\tau} h_t d_{i,t+\tau})' \hat{\Omega}_T^{-1} (T^{-1} \sum_{t=1}^{T-\tau} h_t d_{i,t+\tau}) \sim \chi_1^2 \quad (12)$$

where  $\hat{\Omega}_T$  is the Newey and West (1987) HAC estimator of the asymptotic variance of the  $h_t d_{i,t+\tau}$ . In this application the CPA test is used to assess whether any model under consideration outperforms the random walk model under the squared error metric.

Given that volatility is latent, we follow (Andersen and Bollerslev, 1998) and (Andersen et al., 1999) and we use the realized variance (volatility) as proxy for the true volatility. The realized variance is defined by

$$\sigma_t^2 = \sum_{j=1}^n r_{t,j}^2 \quad (13)$$

where  $r_{t,j}$  is the return in interval  $j$  on day  $t$  and  $n$  is the number of interval in a day. The realized variance we use in this paper are based on the sum of 5-minute intra-day squared returns.<sup>1</sup>

<sup>1</sup>As mentioned in Subsection 2.1, the daily realized variance of all indices are obtained from the Oxford-Man Institute's Realized Library version 0.2.

The ability of the models described in Subsection (2.2) to accurately forecast the 'true' volatility is assessed using three alternative types of measures for forecast comparisons. More specifically, we use the mean absolute error (MAE) and the root mean squared error (RMSE) and the  $R^2$  from the Mincer-Zarnowitz (MZ) regression (Mincer and Zarnowitz, 1969).

$$MAE = \frac{1}{\tau} \sum_{t=T+1}^{T+\tau} |h_t^2 - \sigma_t^2| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{\tau} \sum_{t=T+1}^{T+\tau} (h_t^2 - \sigma_t^2)^2} \quad (15)$$

$$\sigma_t^2 = a_0 + a_1 h_t^2 + \epsilon_t \quad (16)$$

where  $\tau$  is the number of out-of-sample observations,  $h_t^2$  is the GARCH, IV or RV forecast and  $\sigma_t^2$  is the 'true' volatility, which is measured as the sum of intraday squared returns. The MAE measures the average absolute forecast error and by construction does not permit the offsetting effect of over- and under-prediction. The RMSE is a conventional criterion which clearly weights greater forecast errors more heavily than smaller forecast errors in the forecast error penalty. The MZ measures how much of the true volatility is explained by the forecast series. The model with the highest  $R^2$  is preferred.

Finally, in order to examine the relative forecasting performance of the GARCH, IV and RV models, a forecast-encompassing exercise is also performed. To test for such forecast encompassing the following extension of the regression model in equation (16) is considered:

$$\sigma_t^2 = a_0 + a_1 h_{1,t}^{2f} + a_2 h_{2,t}^{2f} + a_3 h_{3,t}^{2f} + \epsilon_t \quad (17)$$

where  $h_{1,t}^{2f}$  refers to the GARCH forecasts,  $h_{2,t}^{2f}$  refers to the IV forecasts and  $h_{3,t}^{2f}$  to the RV forecasts.<sup>2</sup> If the IV forecast model carries no additional information then it is said that the GARCH and/or RV forecast encompasses the IV forecast and the null hypothesis that  $a_2 = 0$  is true. Similarly, if IV encompasses GARCH the null hypothesis that  $a_1 = 0$  is true and so on. To examine whether the IV forecast produced by ARMA models is a better predictor for the future stock index return volatility than the random walk we run the equation (17) twice: first using the forecast of IV indices as  $h_{2,t}^{2f}$  and second using IV following a random walk as  $h_{2,t}^{2f}$ .

---

<sup>2</sup>All the different encompassing regressions considered in this paper can be found on the Appendix A. We examine the relative forecasting performance of GARCH and IV forecast by letting  $a_3 = 0$ , the relative forecasting performance of GARCH and RV forecast by letting  $a_2 = 0$  and so on.

## 3 Empirical results

### 3.1 In-sample results

Here, we report the in-sample estimates for the GARCH and ARMA models for the stock return and IV series respectively. As noted above, the initial in-sample estimation period is from February 2, 2001 to February 23, 2010.

Table 2 reports the parameter estimates, the log-likelihood and the likelihood ratio test<sup>3</sup> of the alternative GARCH models. The period used for the estimations is February 2, 2001 to February 23, 2010. The results are the same for all indices. The coefficient of IV is always significant at the 1% level suggesting that IV contains incremental information useful for explaining the conditional variance. This conclusion is also supported by the log-likelihood and the likelihood ratio test. The log-likelihood of the augmented GARCH specifications is higher than their restricted counterparts. The  $\chi^2$  statistic indicates that the null hypothesis of no additional information content is rejected. The EGARCH-IV model performs best for the S&P500 and Nasdaq100 indices, while the ACGARCH-IV has the highest log-likelihood for the DJIA.

Table 3 summarizes the AR(FI)MA(X) models' coefficients, their p-values and the log-likelihood for all IV indices. The AR(1) and MA(1) terms are statistically significant at the 1% level for all models except the ARFIMAX specification for the VIX index. The coefficients of  $r_t^+$  and  $r_t^-$  are also statistically significant at the 1% level. Moreover, the coefficient of  $r_t^-$  are greater in absolute values than the coefficients of  $r_t^+$  for all models. It is apparent that there is a contemporaneous asymmetric relationship between index returns and IV, which comes in line with the results of Giot (2005b). The difference parameter  $d$  in the ARFIMA models is significant at 1% level and  $d \in (0, 0.5)$  for all series indicating that the processes exhibit long memory. Based on the log-likelihood, the unrestricted ARMA models, those that allow for asymmetry, outperform their restricted counterparts. Overall, within the sample, the ARMAX specification performs best for VIX and VXD indices and the ARIMAX specification for the VXN index.

### 3.2 Out-of-sample results

Our initial out-of-sample period is from February 24, 2010 until the end of the sample, February 28, 2013. As discussed above, the out-of-sample forecasting exercise is conducted recursively, thus,

---

<sup>3</sup> The likelihood ratio test is defined as  $LR = -2(L_r - L_u) \chi^2(m)$  where  $L_u$  is the maximized value of the log likelihood for an unrestricted model (in our case GARCH-IV),  $L_r$  is the maximized value of the log likelihood for a model which has been estimated imposing the constraints (in our case GARCH(1,1) imposing the constraint that  $\theta = 0$ ) and  $m$  is the number of the restrictions.

at each step the in-sample period is expanded and the out-of-sample period reduced. At each step forecasts are obtained as one-step ahead forecasts. The recursive strategy allows use of all available data to generate forecasts but avoids potentially small samples and unstable estimates than can occur with rolling forecasts.

Regarding the predictability of IV itself, we run the CPA test for the squared forecast error. Table 4 reports the p-values for testing the null hypothesis of equal predictive ability between the row and column models. Rejection of the null hypothesis is indicated by the superscripts + and -. A positive (negative) sign indicates that the row (column) model is outperformed by the column (row) model. There are 9 cases (out of 18) in which we reject the null hypothesis that the random walk and the ARMA-type models perform equally well. In these cases, the + sign denotes the ARMA models performs better than the random walk. This suggests that there is a predictable pattern in the dynamics of implied volatility indices, which is in line with the results of Konstantinidi et al. (2008). When the predictive ability of the random walk is tested against the asymmetric ARMA models, the latter performs significantly better than the random walk at the 1% level. The importance of asymmetry found in-sample carries over to the out-of-sample analysis. When the model under consideration is an ARMA model that takes into account the contemporaneous asymmetric effect - ARMAX, ARIMAX, ARFIMAX models - always outperforms not only the random walk, but also the symmetric ARMA models. In these cases, the null hypothesis of equal predictive ability is always rejected at the 1% level.

Table 5 reports the mean absolute error and the root mean square error for the various models. According to the MAE, the RV performs best. More specifically, when RV is forecast by an ARFIMAX model yields the lowest loss for both S&P500 and DJIA indices, while when RV is forecast by an ARIMA specification performs best for the Nasdaq100. Second, the majority of the models that are nested with implied volatility outperform their GARCH counterparts that exclude the implied volatility information. The EGARCH model yields the lowest loss for the S&P500 and Nasdaq100, while the EGARCH-IV performs best for the DJIA index. The IV forecasts perform poorly with the ARMAX and the ARIMAX specification to yield the lowest MAE for S&P500, and DJIA and Nasdaq100, respectively.

When the RMSE is used, once more the RV provides the best forecast. For both S&P500 and DJIA indices, the ARFIMAX specification performs best, while for the Nasdaq100 the ARFIMA yield the lowest loss. Moreover, there is overwhelming evidence of the usefulness of IV in the variance equations as, in all series, a GARCH specification combined with implied volatility outperforms its restricted version. For both the S&P500 and DJIA the EGARCH-IV performs best, while for the Nasdaq100 the GJR-IV provides the best forecast. When the forecasting performance of IV models is assessed, the ARMAX model for the S&P500 index and the ARFIMAX model for the DJIA and

Nasdaq100 indices provides the best forecast. Furthermore, contrary to the MAE results, in many cases ARMA specifications yield lower RMSE than the restricted GARCH specifications.

Table 6 reports the  $R^2$  values from the forecasting regression in equation (16). The model with the highest  $R^2$  is preferred. Examining the results of the MZ test procedure we find that overall the realized volatility forecast by the ARFIMAX model performs best for the S&P500 and DJIA indices and the ARFIMA models is superior for the Nasdaq100. Looking at the GARCH forecasts, the GJR specification embedded with IV performs best for all indices. For all series an unrestricted GARCH specification obtains strictly higher  $R^2$  than its restricted version. The  $R^2$  value rises by about 5% up to 13%. Looking at the IV forecasts, the ARMA-type models which take into account the contemporaneous asymmetric effect - ARMAX, ARIMAX and ARFIMAX - obtain higher  $R^2$  values than the random walk for all cases. The random walk yields marginally higher  $R^2$  than the symmetric ARMA specifications, implying that when the contemporaneous asymmetric effect is considered, the forecast of IV does a better job than the random walk in explaining the variability of the 'true' volatility. Among the IV forecasts, the ARFIMAX specification reports the highest  $R^2$  value across all indices and measures of true volatility.

Furthermore, the relative forecasting performance of the GARCH, IV and RV models is examined so as to identify whether these forecasts contain independent information useful in predicting future volatility and whether the IV forecast through an ARMA-type specification is a better predictor than the random walk, Table 7 summarizes the results of the encompassing regressions described in equation (17) for all indices.<sup>4</sup> There is a remarkable consistency across all indices. First, linear combinations of GARCH, IV and RV forecasts are always more informative than the bivariate mixtures of forecasts and the univariate forecasts presented in Table 6 obtaining strictly higher adjusted  $R^2$ . The value of the adjusted  $R^2$  increases by up to 5.5%. This suggests that GARCH, IV and RV forecasts contain independent information useful in predicting future volatility. A combination of an ACGARCH-IV forecast with the IV and RV forecasts obtained through an asymmetric ARFIMA specification yields the highest adjusted  $R^2$  for all indices. Second, when the relative forecasting performance between GARCH and IV is examined, the GJR-IV combined with the ARMAX model reports the highest adjusted  $R^2$  values for all indices. This is also confirmed looking at the adjusted  $R^2$  which is strictly higher than the  $R^2$  for the univariate regressions in Table 6. However, when GARCH is combined with the RV the ACGARCH-IV and ARFIMAX models perform best for all indices. Third, when the relative forecasting performance between IV and RV is examined, the highest adjusted  $R^2$  is yielded when they are both predicted by an ARFIMAX model.

Regarding the encompassing test, the null hypothesis that the GARCH forecasts encompasses

---

<sup>4</sup>To save space the encompassing regressions for each index are not tabulated here, but can be found on the Appendix B.

the IV forecasts and/or the realized volatility forecasts is rejected for almost all indices. In some cases the coefficient,  $a_1$  of the GARCH forecasts is insignificant indicating that IV and/or RV is more informative than GARCH. Nonetheless, in cases in which one approach dominates the other, the adjusted  $R^2$  is marginally higher than the  $R^2$  of the univariate regressions indicating that combining GARCH, IV and RV improves on the results from either one alone.

## 4 Further international evidence

Since the CBOE launched VIX, IV indices have mushroomed in the European markets. Nevertheless, most studies have focused on the IV for the US markets. To examine the robustness of our results, we repeat the analysis using six European indices and their IV indices over the same sample period, i.e. February 2, 2001 to February 28, 2013. More specifically, the stock indices are the EURO STOXX, CAC40, DAX30, AEX, SMI, FTSE100 and their IV indices are VSTOXX, VCAC, VDAX, VAEX, VSMI, VFTSE100, respectively. The realized variance<sup>5</sup> is used as measure of true volatility.

Similar to the US data, in-sample, the EGARCH-IV models perform best for all indices with the exception of FTSE100 where the GJR-IV has the highest log-likelihood. The inclusion of IV as exogenous variable improves GARCH models' fit. The LR test rejects the null hypothesis that IV does not contain incremental information other than the information contained in GARCH useful for explaining the conditional variance. As far as the IV models are concerned, the ARMAX and ARIMAX specifications perform best for the VSTOXX, VDAX, VAEX and VCAC, VSMI, VFTSE100, respectively. Finally, there is a contemporaneous asymmetric relationship between index returns and IV, which is in line with the US results.<sup>6</sup>

Looking at the forecast of the implied volatility models, Table 8 reports the CPA test results. It can be seen that there are 23 cases (out of 36) in which we reject the null hypothesis of equal predictability. In 22 cases, that is the 61% of the different possible combinations of IV, an ARMA type model performs better than the random walk. This indicates that there is a predictable pattern in the dynamics of implied volatility indices. When an asymmetric ARMA model is competing with the random walk always perform significantly better than the random walk at the 1% level. Moreover, and similar to the US results, the importance of asymmetry is noteworthy also in the context of the European indices. When the model under consideration is an ARMA model that takes into account the contemporaneous asymmetric effect - ARMAX, ARIMAX, ARFIMAX - is superior to the restricted ARMA models. Among the asymmetric ARMA models, we cannot reject

<sup>5</sup>We obtain the daily realized variances from Realized Library version 0.2 of the Oxford-Man Institute of Quantitative Finance Heber et al. (2009). These realized variances are based on the sum of 5-minute intra-day squared returns.

<sup>6</sup>To save space the IV estimation results are not tabulated here, but can be provided on request.

the null hypothesis of equal predictive ability and we conclude that they perform equally well.

In Table 9 the ability of both GARCH and IV models to adequately predict volatility is assessed using the MAE and RMSE. Based on the MAE, the RV performs best yielding the lowest loss. The ARIMA specification provides the best forecast for the DAX, AEX and FTSE indices, while the ARIMAX performs best for the SMI index. As for the STOXX and CAC indices, the ARFIMA and ARFIMAX specifications are superior, respectively. Furthermore, among the GARCH models, the EGARCH-IV provides the best forecast for the CAC, AEX and FTSE100 indices, while the EGARCH performs best for the STOXX and SMI indices. As for the DAX index, the GJR-IV yields the lowest forecast error. Apart from the SMI index, for all indices a GARCH model that embeds IV outperforms its restricted version. As the IV models are concerned, they generally perform poorly. Among the IV models, the ARMAX and ARIMAX specifications yield the lowest MAE. The ARMA models that capture the contemporaneous asymmetric relationship between IV and index returns provide more accurate forecast for the true volatility than their restricted versions and the random walk.

According to the RMSE, when the RV is forecast by the ARFIMA model yields the lowest loss for all indices with the exception of the STOXX index where the asymmetric ARFIMA performs best. The EGARCH-IV provides the best forecast among the GARCH class models for the STOXX, CAC and AEX index, while the GJR-IV and GARCH-IV performs best for the DAX and FTSE100 index, respectively. For the SMI index, similarly to the MAE results, the EGARCH performs best. Among the IV forecasts, the ARFIMAX specification performs best for all indices apart from the SMI in which the ARFIMA specification yields the lowest RMSE.

Table 10 reports the results of the MZ procedure described by equation (16). The model with the highest  $R^2$  is preferred. The results are similar to those obtained using the US data. First, an asymmetric GARCH augmented with IV provides the best forecast for four indices. More specifically, the EGARCH-IV obtains the highest  $R^2$  value STOXX, CAC and SMI indices and the GJR-IV performs best for the FTSE100 index. As for the DAX and AEX indices, the realized volatility forecast by an ARFIMA model is superior. Second, looking at the GARCH forecasts when the GARCH models are augmented with IV perform better than their restricted counterparts with the GJR-IV performing best for the DAX and FTSE100 indices and the EGARCH-IV being superior for all other indices. Third, the asymmetric IV model forecasts have strictly higher  $R^2$  values than the random walk with the exception of the CAC index in which the random walk outperforms the ARMA forecasts.

Finally, to investigate the relative forecasting performance of the models in order to identify whether IV forecasts subsume the information in GARCH and/or RV forecasts, we summarize the encompassing regression results in Table 11.<sup>7</sup> There is a noteworthy consistency across indices.

---

<sup>7</sup>To save space the encompassing regressions for each index are not tabulated here, but can be provided on request.

First, linear combinations of GARCH, IV and RV forecast obtain the highest adjusted  $R^2$  indicating that they all contain independent information useful in predicting future volatility. A combination of an asymmetric GARCH with IV following a RW and RV forecasted by an ARMAX model yields the highest adjusted  $R^2$  for all indices with the exception of FTSE100. In this case, and similarly to the US results, a combination of an ACGARCH-IV with the ARFIMAX specification for both IV and RV is superior. Second, the table shows that a combination of GARCH, IV and RV is more informative than a combination which excludes the information of RV increasing the adjusted  $R^2$  by up to 7% and than a combination which only considers the information of GARCH and RV increasing the adjusted  $R^2$  by up to 4%. Moreover, GARCH carries additional information beyond the one contained in both IV and RV, because the  $R^2$  values of this bivariate regression increase by up to 6% when GARCH information is taken into account. This is consistent to the results obtained for the US indices. All approaches are highly significant implying that they all contain independent information useful in forecasting future volatility. This can also be noticed from the  $R^2$  value which is higher than the  $R^2$  for the univariate regressions presented in Table 10.

## 5 Value-at-risk

The above results suggest that combining forecasts appears to produce an improved forecast performance compared to single model forecasts. Of course, the above results are obtained using statistical measures only. Therefore, in this section, the performance of the forecast encompassing (or combination) regressions are assessed using the value-at-risk (VaR) approach. VaR is a popular technique designed to measure risk as it specifies the potential currency value loss in a portfolio that could occur within a given time frame and with a given probability. Thus, it allows us the ability to assess volatility forecasts using an economic evaluation approach in addition to the above statistical ones. VaR is calculated as the chosen cutoff (or quantile) in the returns distribution. Formally, assuming a normal distribution, the portfolio currency VaR is given as  $VaR = a(N)\sigma_{t+1}V$ , where  $a(N)$  is the appropriate left-hand cut-off of the normal distribution,  $\sigma_{t+1}$  is the one-step ahead volatility forecast and  $V$  is the portfolio's value. Of course, to calculate VaR here we do not need knowledge of  $V$ .

Given the success of the combination forecasts in the above analysis, we focus solely on them here. Thus,  $\sigma_{t+1}$  is the one-step ahead volatility forecast as it is estimated through the combinations of the GARCH, IV and RV model forecasts. In order to evaluate the performance of these forecasts for producing reasonable VaR estimates we examine the models failure rate. That is, the frequency that the actual loss exceeds the estimated VaR. Specifically, two tests are considered. The Kupiec test, which is an unconditional test regarding the number of times the actual loss exceeds to VaR (known as exceptions) and the DQ test, which is a conditional test that examines the nature of the



sequence of exceptions.

Specifically, the Kupiec test (Kupiec, 1995) examines the equality of the empirical failure rate to a specified statistical level. That is, using, for example, the 5% cut-off whether the actual number of exceptions greater than 5%. The dynamic quantile (DQ) test proposed by Engle and Manganelli (2004) argues that in addition to the failure rate, the conditional accuracy of the VaR estimates is important. That is, whether failures are bunched in time. Therefore, they test the joint null hypothesis that the violations should both occur at a specified rate and not be serially correlated. They define the hit sequence:

$$Hit_t = I(r_t < -VaR_t) - a$$

which assumes value  $(1 - a)$  every time the actual return is less than the VaR quantile and  $-a$  otherwise. The expected value of  $Hit_t$  is zero and the hit sequence must be uncorrelated with any past information and have an expected value equal to zero. If the hit sequence satisfies these conditions the hits will not be correlated and the fraction of exception will be correct. The DQ test statistic is computed as

$$DQ = \hat{\beta}' X' X \beta / a(\hat{1} - a)$$

where  $X$  is the vector of explanatory variables and  $\hat{\beta}$  the OLS estimates. The test follows a  $\chi^2$  distribution with degree of freedom equal to the number of parameters.

The 1% VaR results for the encompassing regressions of both the US and European indices reported in Table 12 show that the combinations perform generally poor. Their average failure rate is significantly higher than the 1% VaR level. The combination providing the best VaR measures in terms of achieving the lowest average failure rate is the ACGARCH combined with the realized volatility when the latter is forecast by the ARFIMAX model followed by a combination of ACGARCH with ARFIMA for both IV and RV. Moreover, when GARCH forecasts are combined with the IV forecasts produced by an asymmetric ARMA model have lower average failure rate than when the GARCH models are combined with the IV following a random walk. Overall, the combinations whose explanatory variables take into account for the asymmetric effect perform better obtaining lower average failure rate. In terms of the Kupiec and DQ tests, while the combinations perform poorly, a combination of IV and RV through a ARIMAX model performs best with only 3 indices significant. The other indices reject the null hypotheses of the equality of the number of violations at a specified rate, Kupiec test, and of the non autocorrelation in the sequence of exceptions, DQ test.

Table 13 reports the 5% VaR results for the encompassing regressions of both the US and European indices. Similar to the 1% VaR results, the combination providing the best VaR measures in terms of achieving the lowest average failure rate is the ACGARCH combined with the realized

volatility when the latter is forecast by the ARFIMAX model followed by a combination of ACGARCH with ARFIMA for both IV and RV. In terms of the specification tests, both the ACGARCH model combined with the ARIMAX model for the RV and the ACGARCH model combined with the ARFIMAX model for the RV perform the best, with four indices significant on the Kupiec test and three indices significant on the DQ test. Finally, when GARCH forecasts are combined with the IV forecasts produced by an asymmetric ARMA model have lower average failure rate than when the GARCH models are combined with the IV following a random walk.

## 6 Conclusion

This paper provides a comparative evaluation of the ability of a wide range of GARCH, IV and RV models to forecast stock index return volatility for a number of US and European indices. Recent literature has shown that IV follows a predictable pattern. Therefore, this paper analyzes whether the IV forecasts are good predictors for the stock market volatility. A total of ten GARCH models are considered, GARCH, GJR, EGARCH, CGARCH and ACGARCH model and the encompassing variants of these models including IV as a regressor in the variance equation. Additionally, six ARMA models have been taken into consideration for forecasting IV indices and realized volatility.

The results show that both the IV and RV forecasts contain significant information regarding the future volatility. With regard to the forecasting ability of IV itself, we find that IV forecasts are statistically significant. When the IV model accounts for the contemporaneous asymmetric effect its forecast strictly outperforms the random walk. As for the GARCH models, the inclusion of IV in the GARCH variance equations improves both the in-sample and out-of-sample performance of the GARCH models with an asymmetric GARCH to perform best. Encompassing regressions indicate that a linear combination of GARCH, IV and RV improves the forecasts. Finally, with regard to VaR forecasts, the ACGARCH combined with the realized volatility when the latter is forecasted by the ARFIMAX model is preferred followed by a combination of ACGARCH with ARFIMA for both IV and RV.

To summarize, the results suggest that both IV and RV do contain additional information useful for forecasting future stock market volatility beyond the information contained in the GARCH based models alone. Furthermore, the presence of asymmetric effects is highly important as it significantly improves the performance of GARCH, IV and RV models. Overall, for the US data, a model which combines the information contained in ACGARCH forecast with the information from option markets and information of RV through ARFIMAX models is the most appropriate for predicting future volatility. As for the EU data, a mixture of an asymmetric GARCH, IV and RV forecast through an ARMAX model performs best with the exception of FTSE100 index which is consistent with the US data. It is hoped the results presented here will be interest not only to

academics involved in developing theoretical and empirical models of market behaviour but also to practitioners to whom accurate volatility forecasts are essential.

## References

- Ahoniemi, K. (2006). Modeling and forecasting implied volatility: an econometric analysis of the VIX index. Working paper, Helsinki School of Economics.
- Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- Andersen, T. G., Bollerslev, T., and Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5):457–477.
- Becker, R., Clements, A. E., and White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking and Finance*, 31(8):2535 – 2549.
- Blair, B. J., Poon, S.-H., and Taylor, S. J. (2001). Forecasting S&P100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, 105(1):5 – 26. Forecasting and empirical methods in finance and macroeconomics.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Carr, P. and Wu, L. (2006). A tale of two indices. *Journal of Derivatives*, 13(3):13–29.
- Cheng, X. and Fung, J. K. (2012). The information content of model-free implied volatility. *Journal of Futures Markets*, 32(8):792–806.
- Cheng, X. and Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2):280 – 293.
- Chiras, D. P. and Manaster, S. (1978). The information content of option prices and a test of market efficiency. *Journal of Financial Economics*, 6(2-3):213–234.
- Christensen, B. and Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125 – 150.
- Claessen, H. and Mittnik, S. (2002). Forecasting stock market volatility and the informational efficiency of the DAX-index options market. *European Journal of Finance*, 8(3):302–321.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559 – 583.

- Corrado, C. J. and Miller, T. W. (2005). The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets*, 25(4):339–373.
- Day, T. E. and Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52(1 - 2):267 – 287.
- Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination. In *Handbook of Statistics, Maddala GS, Rao CR(eds)*, volume 14, pages 241–268. North-Holland.
- Dunis, C., Kellard, N. M., and Snaith, S. (2013). Forecasting EUR - USD implied volatility: The case of intraday data. *Journal of Banking and Finance*, 37(12):4943 – 4957.
- Ederington, L. H. and Guan, W. (2002). Measuring implied volatility: Is an average better? which average? *Journal of Futures Markets*, 22(9):811–837.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):pp. 987–1007.
- Engle, R. F. and Lee, G. G. (1993). A permanent and transitory component model of stock return volatility. *University of California at San Diego, Economics Working Paper Series*.
- Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.
- Fernandes, M., Medeiros, M. C., and Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking and Finance*, 40(0):1 – 10.
- Frijns, B., Tallau, C., and Tourani-Rad, A. (2010). The information content of implied volatility: Evidence from Australia. *Journal of Futures Markets*, 30(2):134–155.
- Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.
- Giot, P. (2005a). Implied volatility indexes and daily Value at risk models. *Journal of Derivatives*, 12(4):54–64.
- Giot, P. (2005b). Relationships between implied volatility indices and stock index returns. *Journal of Portfolio Management*, 31(3):92–100.
- Glosten, L. R., Jagannathan, R., and 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):pp. 1779–1801.
- Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). Oxford-man Institute’s Realized Library, Version 0.2.

- Hendry, D. F. and Clements, M. P. (2004). Pooling of forecasts. *Econometrics Journal*, 7(1):1–31.
- Kim, H. H. and Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178 (2):352 – 367.
- Konstantinidi, E., Skiadopoulos, G., and Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and U.S. implied volatility indices. *Journal of Banking and Finance*, 32(11):2401 – 2411.
- Koopman, S. J., Jungbacker, B., and Hol, E. (2005). Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3):445–475.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73.84.
- Lamoureux, C. G. and Lastrapes, W. D. (1993). Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *The Review of Financial Studies*, 6(2):pp. 293–326.
- Latane, H. A. and Rendleman, Richard J., J. (1976). Standard deviations of stock price ratios implied in option prices. *The Journal of Finance*, 31(2):pp. 369–381.
- Martens, M. and Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-a-vis implied volatility. *Journal of Futures Markets*, 24(11):1005–1028.
- Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER Chapters, pages 1–46. National Bureau of Economic Research, Inc.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):pp. 347–370.
- Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, 1:135–196.
- Tsai, W.-C., Chiu, Y.-T., and Wang, Y.-H. (2015). The information content of trading activity and quote changes: Evidence from vix options. *Journal of Futures Markets*, 35(8):715–737.
- Yang, M. and Liu, M. (2012). The forecast power of the volatility index in emerging markets: Evidence from the Taiwan stock market. *International Journal of Economics and Finance*, 4(2):217–231.
- Yu, W. W., Lui, E. C., and Wang, J. W. (2010). The predictive power of the implied volatility of options traded OTC and on exchanges. *Journal of Banking and Finance*, 34(1):1–11.

# Appendix

## A. List of encompassing regressions

Encompassing regressions	
1	$\sigma_t^2 = a_0 + a_1 GARCH_t + a_2 ARMA_t + \epsilon_t$
2	$\sigma_t^2 = a_0 + a_1 GARCH_t + a_3 ARMA_t + \epsilon_t$
3	$\sigma_t^2 = a_0 + a_1 GARCH_t + a_2 RW_t + \epsilon_t$
4	$\sigma_t^2 = a_0 + a_2 ARMA_t + a_3 ARMA_t + \epsilon_t$
5	$\sigma_t^2 = a_0 + a_2 RW_t + a_3 ARMA_t + \epsilon_t$
6	$\sigma_t^2 = a_0 + a_1 GARCH_t + a_2 RW_t + a_3 ARMA_t + \epsilon_t$
7	$\sigma_t^2 = a_0 + a_1 GARCH_t + a_2 ARMA_t + a_3 ARMA_t + \epsilon_t$
8	$\sigma_t^2 = a_0 + a_1 GARCH - IV_t + a_2 ARMA_t + \epsilon_t$
9	$\sigma_t^2 = a_0 + a_1 GARCH - IV_t + a_3 ARMA_t + \epsilon_t$
10	$\sigma_t^2 = a_0 + a_1 GARCH - IV_t + a_2 RW_t + \epsilon_t$
11	$\sigma_t^2 = a_0 + a_1 GARCH - IV_t + a_2 RW_t + a_3 ARMA_t + \epsilon_t$
12	$\sigma_t^2 = a_0 + a_1 GARCH - IV_t + a_2 ARMA_t + a_3 ARMA_t + \epsilon_t$
13	$\sigma_t^2 = a_0 + a_1 GJR_t + a_2 ARMAX_t + \epsilon_t$
14	$\sigma_t^2 = a_0 + a_1 GJR_t + a_3 ARMAX_t + \epsilon_t$
15	$\sigma_t^2 = a_0 + a_1 GJR_t + a_2 RW_t + \epsilon_t$
16	$\sigma_t^2 = a_0 + a_2 ARMAX_t + a_3 ARMAX_t + \epsilon_t$
17	$\sigma_t^2 = a_0 + a_2 RW_t + a_3 ARMAX_t + \epsilon_t$
18	$\sigma_t^2 = a_0 + a_1 GJR_t + a_2 RW_t + a_3 ARMAX_t + \epsilon_t$
19	$\sigma_t^2 = a_0 + a_1 GJR_t + a_2 ARMAX_t + a_3 ARMAX_t + \epsilon_t$
20	$\sigma_t^2 = a_0 + a_1 GJR - IV_t + a_2 ARMAX_t + \epsilon_t$
21	$\sigma_t^2 = a_0 + a_1 GJR - IV_t + a_3 ARMAX_t + \epsilon_t$
22	$\sigma_t^2 = a_0 + a_1 GJR - IV_t + a_2 RW_t + \epsilon_t$
23	$\sigma_t^2 = a_0 + a_1 GJR - IV_t + a_2 RW_t + a_3 ARMAX_t + \epsilon_t$
24	$\sigma_t^2 = a_0 + a_1 GJR - IV_t + a_2 ARMAX_t + a_3 ARMAX_t + \epsilon_t$
25	$\sigma_t^2 = a_0 + a_1 EGARCH_t + a_2 ARMAX_t + \epsilon_t$
26	$\sigma_t^2 = a_0 + a_1 EGARCH_t + a_3 ARMAX_t + \epsilon_t$
27	$\sigma_t^2 = a_0 + a_1 EGARCH_t + a_2 RW_t + \epsilon_t$
28	$\sigma_t^2 = a_0 + a_1 EGARCH_t + a_2 RW_t + a_3 ARMAX_t + \epsilon_t$
29	$\sigma_t^2 = a_0 + a_1 EGARCH_t + a_2 ARMAX_t + a_3 ARMAX_t + \epsilon_t$
30	$\sigma_t^2 = a_0 + a_1 EGARCH - IV_t + a_2 ARMAX_t + \epsilon_t$
31	$\sigma_t^2 = a_0 + a_1 EGARCH - IV_t + a_3 ARMAX_t + \epsilon_t$
32	$\sigma_t^2 = a_0 + a_1 EGARCH - IV_t + a_2 RW_t + \epsilon_t$
33	$\sigma_t^2 = a_0 + a_1 EGARCH - IV_t + a_2 RW_t + a_3 ARMAX_t + \epsilon_t$
34	$\sigma_t^2 = a_0 + a_1 EGARCH - IV_t + a_2 ARMAX_t + a_3 ARMAX_t + \epsilon_t$
35	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 ARIMA_t + \epsilon_t$
36	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 ARFIMA_t + \epsilon_t$
37	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_3 ARIMA_t + \epsilon_t$
38	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_3 ARFIMA_t + \epsilon_t$
39	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 RW_t + \epsilon_t$
40	$\sigma_t^2 = a_0 + a_2 ARIMA_t + a_3 ARIMA_t + \epsilon_t$
41	$\sigma_t^2 = a_0 + a_2 ARFIMA_t + a_3 ARFIMA_t + \epsilon_t$
42	$\sigma_t^2 = a_0 + a_2 RW_t + a_3 ARIMA_t + \epsilon_t$
43	$\sigma_t^2 = a_0 + a_2 RW_t + a_3 ARFIMA_t + \epsilon_t$
44	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 RW_t + a_3 ARIMA_t + \epsilon_t$
45	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 RW_t + a_3 ARFIMA_t + \epsilon_t$
46	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 ARIMA_t + a_3 ARIMA_t + \epsilon_t$
47	$\sigma_t^2 = a_0 + a_1 CGARCH_t + a_2 ARFIMA_t + a_3 ARFIMA_t + \epsilon_t$
48	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 ARIMA_t + \epsilon_t$
49	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 ARFIMA_t + \epsilon_t$
50	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_3 ARIMA_t + \epsilon_t$
51	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_3 ARFIMA_t + \epsilon_t$
52	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 RW_t + \epsilon_t$
53	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 RW_t + a_3 ARIMA_t + \epsilon_t$
54	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 RW_t + a_3 ARFIMA_t + \epsilon_t$
55	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 ARIMA_t + a_3 ARIMA_t + \epsilon_t$
56	$\sigma_t^2 = a_0 + a_1 CGARCH - IV_t + a_2 ARFIMA_t + a_3 ARFIMA_t + \epsilon_t$
57	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 ARIMAX_t + \epsilon_t$
58	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 ARFIMAX_t + \epsilon_t$
59	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_3 ARIMAX_t + \epsilon_t$
60	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_3 ARFIMAX_t + \epsilon_t$
61	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 RW_t + \epsilon_t$
62	$\sigma_t^2 = a_0 + a_2 ARIMAX_t + a_3 ARIMAX_t + \epsilon_t$
63	$\sigma_t^2 = a_0 + a_2 ARFIMAX_t + a_3 ARFIMAX_t + \epsilon_t$
64	$\sigma_t^2 = a_0 + a_2 RW_t + a_3 ARIMAX_t + \epsilon_t$
65	$\sigma_t^2 = a_0 + a_2 RW_t + a_3 ARFIMAX_t + \epsilon_t$
66	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 RW_t + a_3 ARIMAX_t + \epsilon_t$
67	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 RW_t + a_3 ARFIMAX_t + \epsilon_t$
68	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 ARIMAX_t + a_3 ARIMAX_t + \epsilon_t$
69	$\sigma_t^2 = a_0 + a_1 ACGARCH_t + a_2 ARFIMAX_t + a_3 ARFIMAX_t + \epsilon_t$
70	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 ARIMAX_t + \epsilon_t$
71	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 ARFIMAX_t + \epsilon_t$
72	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_3 ARIMAX_t + \epsilon_t$
73	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_3 ARFIMAX_t + \epsilon_t$
74	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 RW_t + \epsilon_t$
75	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 RW_t + a_3 ARIMAX_t + \epsilon_t$
76	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 RW_t + a_3 ARFIMAX_t + \epsilon_t$
77	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 ARIMAX_t + a_3 ARIMAX_t + \epsilon_t$
78	$\sigma_t^2 = a_0 + a_1 ACGARCH - IV_t + a_2 ARFIMAX_t + a_3 ARFIMAX_t + \epsilon_t$

## B. Encompassing regressions result for the US indices

Forecast encompassing regression results for the S&P500 index

S&P500											
Model	$\alpha_0 * 10^{-5}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha dj - R^2$	Model	$\alpha_0 * 10^{-5}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha dj - R^2$
1	-4.56*	0.9854	0.7286*	-	0.3962	27	-3.87*	0.2186***	0.6143*	-	0.4079
	(0.000)	(0.254)	(0.000)				(0.000)	(0.061)	(0.000)		
2	2.15	0.1205***	-	0.7940*	0.4022	28	-3.40*	-0.1573	0.4628*	0.6306*	0.4578
	(0.000)	(0.083)		(0.000)			(0.000)	(0.191)	(0.000)	(0.000)	
3	-4.46*	0.0730	0.7304*	-	0.4060	29	-4.40*	-0.2543**	0.5859*	0.6153*	0.4673
	(0.000)	(0.310)	(0.000)				(0.000)	(0.029)	(0.000)	(0.000)	
4	-3.25*	-	0.4263*	0.5109*	0.4345	30	-3.15*	0.5651*	0.3181*	-	0.4352
	(0.000)		(0.000)	(0.000)			(0.000)	(0.000)	(0.006)		
5	-3.36*	-	0.4474*	0.4823*	0.4412	31	-1.47**	0.4706*	-	0.5162*	0.4657
	(0.000)		(0.000)	(0.000)			(0.027)	(0.000)		(0.000)	
6	-3.68*	-0.2234*	0.5568*	0.6901*	0.4462	32	-1.61*	0.8253*	0.0545	-	0.4297
	(0.000)	(0.005)	(0.000)	(0.000)			(0.000)	(0.300)	(0.683)		
7	-3.65*	-0.2133*	0.5394*	0.6095*	0.4388	33	-1.27	0.5103*	-0.0397	0.5185*	0.4651
	(0.000)	(0.009)	(0.000)	(0.000)			(0.178)	(0.005)	(0.761)	(0.000)	
8	-6.32*	-1.1479*	1.7076*	-	0.4020	34	-2.66*	0.2772**	0.2091***	0.4973*	0.4674
	(0.000)	(0.003)	(0.000)				(0.004)	(0.026)	(0.065)	(0.000)	
9	-2.42*	0.4916*	-	0.5611*	0.4345	35	-4.42*	0.0692	0.7328*	-	0.3882
	(0.002)	(0.000)		(0.000)			(0.000)	(0.363)	(0.000)		
10	-6.99*	-2.4933*	2.7074*	-	0.4333	36	-4.75*	0.0669	0.7508*	-	0.3920
	(0.000)	(0.000)	(0.000)				(0.000)	(0.370)	(0.000)		
11	-5.11*	-1.5286*	1.7061*	0.3639*	0.4488	37	1.58**	-0.0287	-	0.8872*	0.4002
	(0.000)	(0.001)	(0.000)	(0.000)			(0.020)	(0.717)		(0.000)	
12	-2.86*	0.2478	0.2144	0.5334*	0.4340	38	3.17	0.0657	-	0.8845*	0.4179
	(0.009)	(0.567)	(0.568)	(0.000)			(0.963)	(0.318)		(0.000)	
13	-4.10*	0.1939**	0.6288*	-	0.4239	39	-4.54*	0.0424	0.7560*	-	0.4055
	(0.000)	(0.011)	(0.000)				(0.000)	(0.553)	(0.000)		
14	-1.46	0.2450*	-	0.6745*	0.4448	40	-2.21**	-	0.8864*	0.5088*	0.4247
	(0.817)	(0.000)		(0.000)			(0.013)		(0.000)	(0.000)	
15	-3.08*	0.2590*	0.5287*	-	0.4141	41	-2.86*	-	0.3544*	0.6050*	0.4376
	(0.000)	(0.007)	(0.000)				(0.001)		(0.000)	(0.000)	
16	-3.58*	-	0.4206*	0.5520*	0.4646	42	-2.66*	-	0.4424*	0.4460*	0.4341
	(0.000)		(0.000)	(0.000)			(0.002)		(0.000)	(0.000)	
17	-2.93*	-	0.3615*	0.5940*	0.4573	43	-3.04*	-	0.3925*	0.5543*	0.4450
	(0.000)		(0.000)	(0.000)			(0.000)		(0.000)	(0.000)	
18	-2.88*	0.0101	0.3540*	0.5901*	0.4566	44	-2.70*	-3180*	0.5612*	0.6365*	0.4437
	(0.002)	(0.900)	(0.000)	(0.000)			(0.002)	(0.000)	(0.000)	(0.000)	
19	-3.89*	-0.0557	0.4630*	0.5734*	0.4643	45	-3.34*	-0.1976*	0.4959*	0.6420*	0.4492
	(0.000)	(0.486)	(0.000)	(0.000)			(0.000)	(0.009)	(0.000)	(0.000)	
20	-2.63*	0.5151*	0.3222*	-	0.4477	46	-2.45*	-0.3258*	0.5267*	0.6876*	0.4341
	(0.004)	(0.000)	(0.000)				(0.006)	(0.000)	(0.000)	(0.000)	
21	-6.71	0.4427*	-	0.4725*	0.4620	47	-3.24*	-0.1755**	0.4558*	0.6768*	0.4406
	(0.289)	(0.000)		(0.000)			(0.000)	(0.024)	(0.000)	(0.000)	
22	-1.66***	0.6071*	0.9057*	-	0.4419	48	-4.41*	0.0432	0.3893*	-	0.3877
	(0.066)	(0.000)	(0.022)				(0.000)	(0.721)	(0.000)		
23	-1.79*	0.3121*	0.1584***	0.4585*	0.4635	49	-5.00*	-0.00063	0.8108*	-	0.3913
	(0.044)	(0.002)	(0.073)	(0.000)			(0.000)	(0.996)	(0.000)		
24	-2.69*	0.2316**	0.2708*	0.4452*	0.4681	50	-6.46	0.2376*	-	0.6224*	0.4094
	(0.003)	(0.016)	(0.002)	(0.000)			(0.333)	(0.001)		(0.000)	
25	-4.91*	0.1215	0.7308*	-	0.4198	51	-4.04	0.2281*	-	0.7104*	0.4277
	(0.000)	(0.278)	(0.000)				(0.542)	(0.000)		(0.000)	
26	-5.66	0.2894*	-	0.6956*	0.4426	52	-5.20*	-0.1051	0.8966*	-	0.4059
	(0.390)	(0.000)		(0.000)			(0.000)	(0.352)	(0.000)		

Note: Entries are the estimated coefficients, their p-values in parentheses and the adjusted  $R^2$  values from the encompassing regression described in equation (17). \*, \*\*, \*\*\* denote significance at the 1% level. A significant p-value indicates that the forecast under consideration is not encompassed by the alternative model, † denotes the highest adjusted  $R^2$  value.



Forecast encompassing regression results for the DJIA index

DJIA																	
Model	$\alpha_0 * 10^{-5}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha dj - R^2$	Model	$\alpha_0 * 10^{-5}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha dj - R^2$	Model	$\alpha_0 * 10^{-5}$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha dj - R^2$
1	-3.76*	0.0345	0.8519*	-	0.3534	27	-3.15*	0.1090	0.7046*	-	0.3668	53	-3.10*	-0.3080***	0.7840*	0.4289*	0.3894
2	0.140	0.1744**	-	0.7430*	0.3555	28	-3.08*	-0.3314***	0.5080*	0.7076*	0.4336	54	-3.27*	-0.2254	0.6652*	0.5200*	0.4018
3	-3.67*	0.0202	0.8545*	-	0.3648	29	-3.52*	-0.2609**	0.5570*	0.7075*	0.4367	55	-2.11**	-0.0652	0.5194*	0.4688*	0.3767
4	-2.91*	-	0.4944*	0.4848*	0.3905	30	-2.17**	0.7507*	0.2534***	-	0.3899	56	-2.44**	0.0077	0.4170**	0.5599*	0.3923
5	-3.01*	-	0.5201*	0.4545*	0.3981	31	-1.37**	0.4445*	-	0.6025*	0.4344	57	-4.09*	-0.0431	0.9287*	-	0.3652
6	-3.36*	-0.2745*	0.6703*	0.5511*	0.4041	32	-1.54***	0.8593*	0.1240	-	0.3874	58	-4.56*	-0.0405	0.9561*	-	0.3703
7	-3.35*	-0.2632*	0.6505*	0.5739*	0.3956	33	-1.79**	0.3317**	0.1076	0.6016*	0.4341	59	1.20***	-0.0660	-	0.9505*	0.3971
8	-4.11*	-0.3138	1.1277*	-	0.3537	34	-2.27*	0.2450***	0.2038	0.5964*	0.4356	60	-2.64	0.0172	-	0.9721*	0.4348
9	-2.43*	0.6034*	0.6034*	0.5195*	0.3949	35	-3.82*	-0.0656	0.9278*	-	0.3487	61	-3.74*	-0.0175	0.8860*	-	0.3648
10	-4.91*	-2.1207*	2.4926*	-	0.3804	36	-4.10*	-0.0690	0.9468*	-	0.3518	62	-1.65**	-	0.3592*	0.5992*	0.4121
11	-3.51*	-0.7290	1.1196**	0.4000*	0.3987	37	1.71*	-0.0646	-	0.8982*	0.3509	63	-2.40*	-	0.3033*	0.7376*	0.4476
12	-1.63*	1.2701*	-0.5851	0.5895*	0.3958	38	0.382	0.0360	-	0.8880*	0.3687	64	-1.51***	-	0.3485*	0.6010*	0.4117
13	-2.99*	0.2249*	0.6641*	-	0.3736	39	-3.81*	-0.0619	0.9224*	-	0.3655	65	-1.96**	-	0.2581*	0.7625*	0.4446
14	-0.328	0.1937*	-	0.7649*	0.4190	40	-1.01**	-	0.4631*	0.4644*	0.3774	66	-1.63**	-0.2096*	0.4512*	0.6849*	0.4181
15	-2.54*	0.2374*	0.6261*	-	0.3714	41	-2.47*	-	0.4243*	0.5599*	0.3931	67	-2.16*	-0.1218***	0.3426*	0.7844*	0.4465
16	-2.83*	-	0.3793*	0.6491*	0.4343	42	-2.30*	-	0.5242*	0.4040*	0.3872	68	-1.89**	-0.2300*	0.4835*	0.6840*	0.4198
17	-2.53*	-	0.3519*	0.6597*	0.4319	43	-2.63*	-	0.4647*	0.5134*	0.4009	69	-2.73*	-0.1490**	0.4123*	0.7648*	0.4508†
18	-2.77*	-0.0583	0.3966*	0.6793*	0.4315	44	-2.30*	-0.2544*	0.6252*	0.5306*	0.3964	70	-3.27*	0.2103	0.6913*	-	0.3665
19	-3.22*	-0.0851	0.4463*	0.6786*	0.4342	45	-2.84*	-0.1803*	0.5739*	0.5693*	0.4063	71	-3.85*	0.1590	0.7658*	-	0.3710
20	-1.89**	0.5713*	0.3372*	-	0.3966	46	-2.08**	-0.2670*	0.5889*	0.5821*	0.3871	72	0.092	0.2805*	-	0.6654*	0.4056
21	-0.74	0.3737*	-	0.6053*	0.4307	47	-2.81*	-0.1836*	0.5470*	0.6052*	0.3979	73	-8.72	0.2238*	-	0.8006*	0.4422
22	-1.49***	0.6017*	0.2887*	-	0.3941	48	-3.41*	0.0801	0.7961*	-	0.3482	74	-3.18*	0.1636	0.7181*	-	0.3656
23	-1.94**	0.1949***	0.2189**	0.5889*	0.4335	49	-3.83*	0.0292	0.8588*	-	0.3510	75	-1.86**	-0.1271	0.4550*	0.6154*	0.4114
24	-2.30*	0.1658	0.2670*	0.5845*	0.4353	50	0.133	0.3922*	-	0.5349*	0.3697	76	-2.04**	-0.0269	0.2820***	0.7640*	0.4439
25	-3.59*	0.1917	0.7388*	-	0.3696	51	-6.59	0.3636*	-	0.6254*	0.3884	77	-1.91*	-0.0907	0.4342*	0.6119*	0.4116
26	-0.65	0.2355*	-	0.7729*	0.4182	52	-4.16*	-0.1601	1.0175*	-	0.3656	78	-2.79*	-0.1131	0.4036*	0.7491*	0.4473

Note: Entries are the estimated coefficients, their p-values in parentheses and the adjusted  $R^2$  values from the encompassing regression described in equation (17). \*, \*\*, \*\*\* denote significance at the 1% level. A significant p-value indicates that the forecast under consideration is not encompassed by the alternative model, † denotes the highest adjusted  $R^2$  value.



Table 1: Summary statistics for the daily stock index returns and implied volatility indices

	A) Stock index returns			B) IV indices		
	S&P500	DJIA	Nasdaq100	VIX	VXD	VXN
Mean	3.80E-05	8.47E-05	3.37E-05	0.0137	0.0127	0.0179
Median	0.0006	0.0004	0.0008	0.0123	0.0115	0.0150
Maximum	0.1095	0.1051	0.1185	0.0509	0.0470	0.0508
Minimum	-0.0947	-0.0820	-0.1111	0.0062	0.0058	0.0079
Std. Dev.	0.0134	0.0125	0.0180	0.0060	0.0056	0.0082
Skewness	-0.1704	0.0328	0.0592	1.9049	1.8183	1.2459
Kurtosis	11.049	10.778	7.3248	8.4758	7.8965	3.7623
Jarque-Bera	8223.4	7666.2	2371.7	5640.3	4715.2	860.67
p-value	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
ADF (p-value)				0.0051*	0.0091*	0.0113**

Notes: Entries report the summary statistics of the daily stock returns (A) and the daily implied volatility indices (B) for period February 2, 2001 to February 28, 2013. In the last two rows, the p-values of the Jarque-Bera test for normality and the Augmented Dickey-Fuller (ADF) test for unit root are reported. \* and \*\* denote rejection of the null hypothesis at the 1% and 5% level, respectively.

Table 2: Estimation models of the GARCH family

	GARCH	GARCH - IV	GJR	GJR - IV	EGARCH	EGARCH - IV	CGARCH	CGARCH - IV	ACGARCH	ACGARCH - IV
S&P500										
$a_0$	$1.08 * 10^{-6} **$ (0.020)	$-1.37 * 10^{-5} **$ (0.012)	$9.06 * 10^{-7} *$ (0.000)	$4.37 * 10^{-7}$ (0.269)	$-0.18015 *$ (0.000)	$-0.00584$ (0.919)	$0.00016 ***$ (0.068)	$-8.43 * 10^{-6}$ (0.418)	$0.00015 *$ (0.009)	$-1.26 * 10^{-5} *$ (0.007)
$a_1$	$0.07416 *$ (0.000)	$-0.04184 ***$ (0.089)	$-0.02695 **$ (0.012)	$-0.05644 *$ (0.000)	$0.08435 *$ (0.000)	$0.00884$ (0.807)	$0.00244$ (0.355)	$-0.08964 *$ (0.000)	$-0.14559 *$ (0.000)	$-0.01220$ (0.993)
$\beta_1$	$0.91892 *$ (0.000)	$-0.06396$ (0.799)	$0.95343 *$ (0.000)	$0.90724 *$ (0.000)	$0.98754 *$ (0.000)	$0.87468 *$ (0.000)	$-0.99442 *$ (0.000)	$0.19152$ (0.275)	$0.66490 *$ (0.000)	$0.08828$ (0.946)
$\gamma$			$0.13144 *$ (0.000)	$0.18280 *$ (0.000)	$-0.11406 *$ (0.000)	$-0.19229 *$ (0.000)			$0.06728$ (0.216)	$0.05027$ (0.361)
$\rho$							$0.99349 *$ (0.000)	$0.97684 *$ (0.000)	$0.98790 *$ (0.000)	$0.03837$ (0.977)
$\phi$							$0.07334 *$ (0.000)	$0.03265 *$ (0.000)	$0.10462 *$ (0.000)	$-0.04816$ (0.971)
$\theta$	$0.86531 *$ (0.000)			$0.03889 *$ (0.000)	$0.13150 *$ (0.000)	$0.13150 *$ (0.000)		$0.76574 *$ (0.000)		$0.70726 **$ (0.026)
Log - L	7133.993	7177.89	7185.029	7199.599	7178.78	7202.876	7135.571	7198.065	7144.343	7179.536
$\chi^2$	87.79	—	29.14	—	48.19	—	126.16	—	70.39	—
DJIA										
$a_0$	$1.10 * 10^{-6} *$ (0.009)	$-8.73 * 10^{-6} ***$ (0.081)	$9.91 * 10^{-7} *$ (0.000)	$-3.45 * 10^{-7}$ (0.483)	$-0.19593 *$ (0.000)	$-0.06522$ (0.233)	$0.000142 **$ (0.010)	$-3.54 * 10^{-6}$ (0.633)	$0.00013 *$ (0.009)	$-4.53 * 10^{-6} *$ (0.000)
$a_1$	$0.07787 *$ (0.000)	$-0.04274 *$ (0.000)	$-0.07787 *$ (0.229)	$-0.04959 *$ (0.000)	$0.09728 *$ (0.000)	$0.03378$ (0.253)	$-0.10033 *$ (0.000)	$-0.09671 *$ (0.000)	$-0.17666 *$ (0.000)	$-0.12228 **$ (0.030)
$\beta_1$	$0.91451 *$ (0.000)	$-0.08851$ (0.671)	$0.94073 *$ (0.000)	$0.88927 *$ (0.000)	$0.98704 *$ (0.000)	$0.89445 *$ (0.000)	$0.54646 *$ (0.004)	$0.05205$ (0.758)	$0.59917 *$ (0.000)	$0.15721$ (0.393)
$\gamma$			$0.13532 *$ (0.000)	$0.18267 *$ (0.000)	$-0.11315 *$ (0.000)	$-0.17441 *$ (0.000)			$0.11191 *$ (0.005)	$0.06680 **$ (0.042)
$\rho$							$0.98804 *$ (0.000)	$0.96711 *$ (0.000)	$0.98839 *$ (0.000)	$0.96339 *$ (0.000)
$\phi$							$0.1010 *$ (0.000)	$0.02895 *$ (0.003)	$0.09750 *$ (0.000)	$0.03257 *$ (0.000)
$\theta$	$0.87797 *$ (0.000)			$0.04918 *$ (0.000)	$0.10574 *$ (0.000)	$0.10574 *$ (0.000)		$0.81980 *$ (0.000)		$0.74265 *$ (0.000)
Log - L	7233.753	7281.437	7278.800	7296.655	7279.334	7301.833	7242.652	7297.520	7248.263	7302.492
$\chi^2$	95.37	—	35.71	—	45.00	—	109.74	—	108.46	—
Nasdaq100										
$a_0$	$1.14 * 10^{-6} **$ (0.029)	$-1.66 * 10^{-5} ***$ (0.073)	$1.04 * 10^{-6} **$ (0.013)	$-2.63 * 10^{-7}$ (0.843)	$-0.12253 *$ (0.000)	$0.06063$ (0.432)	$-0.00219$ (0.942)	$-1.26 * 10^{-5}$ (0.359)	$0.000299$ (0.109)	$-1.57 * 10^{-5}$ (0.282)
$a_1$	$0.05582 *$ (0.000)	$-0.02184$ (0.442)	$-0.00811$ (0.363)	$-0.03667 **$ (0.025)	$0.07336 *$ (0.000)	$-0.01776$ (0.624)	$-0.01046$ (0.439)	$-0.04893$ (0.104)	$-0.08887 *$ (0.005)	$-0.06980 **$ (0.044)
$\beta_1$	$0.94013 *$ (0.000)	$-0.04489$ (0.887)	$0.95980 *$ (0.000)	$0.83725 *$ (0.000)	$0.99234 *$ (0.000)	$0.76755 *$ (0.000)	$0.9099 *$ (0.000)	$0.09177$ (0.711)	$0.35552$ (0.322)	$0.42994 **$ (0.032)
$\gamma$			$0.08657 *$ (0.000)	$0.14345 *$ (0.000)	$-0.07811 *$ (0.000)	$-0.76755 *$ (0.000)			$0.04185$ (0.348)	$0.06627 ***$ (0.076)
$\rho$							$0.99999 *$ (0.000)	$0.97961 *$ (0.000)	$0.99516 *$ (0.000)	$0.98051 *$ (0.000)
$\phi$							$0.05269 *$ (0.000)	$0.01848 **$ (0.032)	$0.06868 *$ (0.000)	$0.02152 **$ (0.021)
$\theta$								$0.82964 *$ (0.000)		$0.54522 *$ (0.005)
Log - L	6273.628	6315.092	6300.433	6317.492	6303.448	6328.670	6272.703	6323.778	6279.973	6325.611
$\chi^2$	82.93	—	34.12	—	50.44	—	102.15	—	91.28	—

Note: Entries report results of the alternative GARCH models as described in equations (2) - (6). The p-values of the estimated coefficients are in parentheses, the log-likelihood (Log-L) and  $\chi^2$  for testing the restricted GARCH specifications, where  $\theta = 0$ , against the augmented GARCH specifications, where  $\theta \neq 0$  are reported. \*, \*\* and \*\*\* denote significance at the 1%, 5% and 10% level, respectively.

Table 3: Estimation output of time series models for implied volatility prediction

	ARMA	ARMAX	ARIMA	ARIMAX	ARFIMA	ARFIMAX
VIX						
$c_0$	0.01383* (0.000)	0.01374* (0.000)	$-1.85 * 10^{-7}$ (0.991)	$-4.64 * 10^{-7}$ (0.982)	0.01457 (0.440)	0.01454 (0.436)
$AR(1)$	0.98932* (0.000)	0.98813* (0.000)	0.56069* (0.000)	-0.44888* (0.000)	0.89105* (0.000)	0.71781* (0.000)
$MA(1)$	-0.14282* (0.000)	0.15798* (0.000)	-0.70967* (0.000)	0.59274* (0.000)	-0.59004* (0.000)	-0.06291 (0.124)
$d$					0.49314* (0.000)	0.49651* (0.000)
$r_t^+$		-0.02337* (0.000)		-0.02331* (0.000)		-0.02323* (0.000)
$r_t^-$		-0.04256* (0.000)		-0.04160* (0.000)		-0.04284* (0.000)
$Log - L$	12332.11	12914.04	12338.69	12910.83	12317.91	12874.81
VXD						
$c_0$	0.01291* (0.000)	0.01288* (0.000)	$-2.81 * 10^{-7}$ (0.985)	$-4.99 * 10^{-7}$ (0.978)	0.01350 (0.447)	0.01323 (0.347)
$AR(1)$	0.99098* (0.000)	0.99029* (0.000)	0.43836* (0.000)	-0.67812* (0.000)	0.91920* (0.000)	0.81770* (0.000)
$MA(1)$	-0.18672* (0.000)	0.05046* (0.000)	-0.61659* (0.000)	0.74711* (0.000)	-0.65026* (0.000)	-0.28437* (0.000)
$d$					0.49199* (0.000)	0.49335* (0.000)
$r_t^+$		-0.02607* (0.000)		-0.02602* (0.000)		-0.02566* (0.000)
$r_t^-$		-0.03674* (0.000)		-0.03557* (0.000)		-0.03733* (0.000)
$Log - L$	12590.77	13097.63	12590.79	13095.62	12579.71	13064.72
VXN						
$c_0$	0.01978* (0.000)	0.01973* (0.000)	$-9.27 * 10^{-6}$ (0.60)	$-9.57 * 10^{-6}$ (0.655)	0.02128 (0.320)	0.02127 (0.324)
$AR(1)$	0.99405* (0.000)	0.99373* (0.000)	0.73707* (0.000)	-0.63943* (0.000)	0.86512* (0.000)	0.79548* (0.000)
$MA(1)$	-0.05258* (0.000)	0.10309* (0.000)	-0.80766* (0.000)	0.74476* (0.000)	-0.45675* (0.000)	-0.21329* (0.000)
$d$					0.49423* (0.000)	0.49583* (0.000)
$r_t^+$		-0.01351* (0.000)		-0.01367* (0.000)		-0.01352* (0.000)
$r_t^-$		-0.02111* (0.000)		-0.02041* (0.000)		-0.02118* (0.000)
$Log - L$	12418.49	12702.3	12421.5	12707.57	129397.01	12666.19

Note: Entries report results of the alternative implied volatility models as described in equations (7) - (9). The p-values of the estimated coefficients are in parentheses. \* denotes significance at the 1% level.

Table 4: Conditional Giacomini-White test for the US indices

	ARMA(1,1)	ARMAX(1,1)	ARIMA(1,1)	ARIMAX(1,1)	ARFIMA(1,1)	ARFIMAX(1,1)
<u>VIX</u>						
Random walk	0.392	0.000(+)	0.306	0.000(+)	0.580	0.000(+)
ARMA(1,1)	-	0.000(+)	0.250	0.000(+)	0.900	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.303	0.000(-)	0.964
ARIMA(1,1)			-	0.000(+)	0.110	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.561
ARFIMA(1,1)					-	0.000(+)
<u>VXD</u>						
Random walk	0.327	0.000(+)	0.314	0.000(+)	0.494	0.000(+)
ARMA(1,1)	-	0.000(+)	0.498	0.000(+)	0.892	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.137	0.000(-)	0.499
ARIMA(1,1)			-	0.000(+)	0.052(-)	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.552
ARFIMA(1,1)					-	0.000(+)
<u>VXN</u>						
Random walk	0.569	0.000(+)	0.292	0.000(+)	0.882	0.000(+)
ARMA(1,1)	-	0.000(+)	0.272	0.000(+)	0.655	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.406	0.000(-)	0.904
ARIMA(1,1)			-	0.000(+)	0.191	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.610
ARFIMA(1,1)					-	0.000(+)

Note: The p-values of the conditional Giacomini-White test are reported. The null hypothesis that the row model and column model perform equally well is tested in terms of squared forecast error. The superscripts + and - indicate rejection of the null hypothesis, with a positive (negative) sign denoting that the row (column) model is outperformed by the column (row) model.

Table 5: MAE and RMSE

		MAE			RMSE		
		<i>S&amp;P500</i>	<i>DJIA</i>	<i>Nasdaq100</i>	<i>S&amp;P500</i>	<i>DJIA</i>	<i>Nasdaq100</i>
GARCH	GARCH	0.0796	0.0710	0.1016	0.1461	0.1365	0.1551
	GARCH-IV	0.0777	0.0667	0.0974	0.1367	0.1288	0.1414
	GJR	0.0763	0.0689	0.1004	0.1429	0.1337	0.1532
	GJR-IV	0.0740	0.0661	0.0962	0.1332	0.1254	0.1409
	EGARCH	0.0679	0.0637	0.0933	0.1341	0.1289	0.1459
	EGARCH-IV	0.0725	0.0634	0.0950	0.1313	0.1241	0.1412
	CGARCH	0.0808	0.0729	0.1036	0.1473	0.1458	0.1582
	CGARCH-IV	0.0782	0.0664	0.0976	0.1449	0.1307	0.1500
	ACGARCH	0.0807	0.0719	0.1035	0.1593	0.1439	0.1632
	ACGARCH-IV	0.0778	0.0659	0.0979	0.1357	0.1288	0.1517
IV	ARMA	0.1180	0.0945	0.1453	0.1591	0.1396	0.1807
	ARMAX	0.1167†	0.0940	0.1449	0.1562†	0.1380	0.1793
	ARIMA	0.1183	0.0946	0.1442	0.1600	0.1401	0.1798
	ARIMAX	0.1169	0.0939†	0.1440†	0.1569	0.1384	0.1788
	ARFIMA	0.1184	0.0947	0.1455	0.1593	0.1398	0.1801
	ARFIMAX	0.1171	0.0942	0.1451	0.1558	0.1377†	0.1786†
	Random walk	0.1178	0.0942	0.1445	0.1588	0.1388	0.1802
	RV	ARMA	0.0635	0.0646	0.0379	0.1309	0.1277
ARMAX	0.0609	0.0599	0.0389	0.1271	0.1212	0.1078	
ARIMA	0.0620	0.0626	0.0354*	0.1315	0.1283	0.1064	
ARIMAX	0.0600	0.0591	0.0364	0.1287	0.1230	0.1072	
ARFIMA	0.0627	0.0633	0.0362	0.1286	0.1257	0.1045*	
ARFIMAX	0.0591*	0.0585 *	0.0370	0.1236*	0.1186*	0.1052	

Note: The mean absolute forecast error (MAE) and the root mean squared forecast error (RMSE) defined in equations (14) and (15), respectively, of GARCH, IV and RV models are reported. All numbers are multiplied by  $10^3$ . \* denotes the lowest forecast error. † denotes the lowest forecast error among the IV models.

Table 7: Forecast encompassing regression results

S&P500		DJIA		Nasdaq100		
Model	$adj - R^2$	Model	$adj - R^2$	Model	$adj - R^2$	
GARCH & IV	GJR-IV & ARX	0.4477	GJR-IV & ARX	0.3966	GJR-IV & ARX	0.3412
GARCH & RV	ACG-IV & ARFX	0.4731	ACG-IV & ARFX	0.4422	ACG-IV & ARFX	0.3612
IV & RV	ARFX & ARFX	0.4807	ARFX & ARFX	0.4476	ARFX & ARFX	0.3701
GARCH & IV & RV	ACG & ARFX & ARFX	0.4868	ACG & ARFX & ARFX	0.4508	ACG & ARFX & ARFX	0.3809

Note: Entries are the encompassing regressions that yield the highest adjusted  $R^2$  values from the encompassing regression described in equation (17).

Table 6: Out-of-sample predictive power of daily volatility forecasts

Models		$adj - R^2$		
		<i>S&amp;P500</i>	<i>DJIA</i>	<i>Nasdaq100</i>
GARCH	GARCH	0.3207	0.2771	0.1946
	GARCH-IV	0.3786	0.3447	0.2970
	GJR	0.3849	0.3368	0.2416
	GJR-IV	0.4387	0.3888	0.3406
	EGARCH	0.3797	0.3355	0.2018
	EGARCH-IV	0.4303	0.3877	0.3197
	CGARCH	0.3135	0.2163	0.1985
	CGARCH-IV	0.3613	0.3293	0.2852
	ACGARCH	0.2567	0.2317	0.1617
	ACGARCH-IV	0.3900	0.3593	0.2943
IV	ARMA	0.3960	0.3541	0.3072
	ARMAX	0.4197	0.3684	0.3179
	ARIMA	0.3884	0.3488	0.3017
	ARIMAX	0.4161	0.3657	0.3163
	ARFIMA	0.3921	0.3518	0.3074
	ARFIMAX	0.4206	0.3709	0.3196
	RW: $IV_{t-1}$	0.4060	0.3656	0.3106
RV	ARMA	0.4006	0.3520	0.3296
	ARMAX	0.4323	0.4123	0.3196
	ARIMA	0.4009	0.3511	0.3371
	ARIMAX	0.4246	0.3971	0.3302
	ARFIMA	0.4179	0.3692	0.3514*
	ARFIMAX	0.4617*	0.4355*	0.3444

Note: Entries are the adjusted  $R^2$  values from the Mincer-Zarnowitz regression described in equation (16). \* denotes the highest adjusted  $R^2$  value.



Table 8: Conditional Giacomini-White test for the European indices

	ARMA(1,1)	ARMAX(1,1)	ARIMA(1,1)	ARIMAX(1,1)	ARFIMA(1,1)	ARFIMAX(1,1)
<u>VST</u>						
Random walk	0.533	0.000(+)	0.488	0.000(+)	0.620	0.000(+)
ARMA(1,1)	-	0.000(+)	0.553	0.000(+)	0.667	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.696	0.000(-)	0.819
ARIMA(1,1)			-	0.000(+)	0.232	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.959
ARFIMA(1,1)					-	0.000(+)
<u>VCC</u>						
Random walk	0.397	0.000(+)	0.715	0.000(+)	0.428	0.001(+)
ARMA(1,1)	-	0.000(+)	0.086(+)	0.000(+)	0.607	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.079(+)	0.000(-)	0.834
ARIMA(1,1)			-	0.000(+)	0.306	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.278
ARFIMA(1,1)					-	0.000(+)
<u>VDX</u>						
Random walk	0.025(+)	0.000(+)	0.062(+)	0.000(+)	0.485	0.000(+)
ARMA(1,1)	-	0.000(+)	0.080(-)	0.000(+)	0.356	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.128	0.000(-)	0.805
ARIMA(1,1)			-	0.000(+)	0.398	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.752
ARFIMA(1,1)					-	0.000(+)
<u>VAX</u>						
Random walk	0.112	0.000(+)	0.120	0.000(+)	0.578	0.000(+)
ARMA(1,1)	-	0.000(+)	0.094(-)	0.000(+)	0.194	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.176	0.000(-)	0.203
ARIMA(1,1)			-	0.000(+)	0.227	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.247
ARFIMA(1,1)					-	0.000(+)
<u>VSM</u>						
Random walk	0.014(+)	0.001(+)	0.091(+)	0.029(+)	0.655	0.001(+)
ARMA(1,1)	-	0.040(+)	0.288	0.297	0.397	0.013(+)
ARMAX(1,1)		-	0.014(-)	0.193	0.079(-)	0.849
ARIMA(1,1)			-	0.144	0.492	0.028(+)
ARIMAX(1,1)				-	0.268	0.631
ARFIMA(1,1)					-	0.007(+)
<u>VFT</u>						
Random walk	0.261	0.000(+)	0.243	0.000(+)	0.071(-)	0.000(+)
ARMA(1,1)	-	0.000(+)	0.283	0.000(+)	0.007(-)	0.000(+)
ARMAX(1,1)		-	0.000(-)	0.586	0.000(-)	0.016(-)
ARIMA(1,1)			-	0.000(+)	0.121	0.000(+)
ARIMAX(1,1)				-	0.000(-)	0.272
ARFIMA(1,1)					-	0.000(+)

Note: The p-values of the conditional Giacomini-White test are reported. The null hypothesis that the row model and column model perform equally well is tested in terms of squared forecast error. The superscripts + and - indicate rejection of the null hypothesis, with a positive (negative) sign denoting that the row (column) model is outperformed by the column (row) model.

Table 9: MAE and RMSE for the European indices

	MAE										RMSE									
	STOXX	CAC	DAX	AEX	SMI	FTSE100	STOXX	CAC	DAX	AEX	SMI	FTSE100	STOXX	CAC	DAX	AEX	SMI	FTSE100		
GARCH	GARCH	0.1206	0.1236	0.0994	0.0837	0.0567	0.0704	0.2019	0.1685	0.1365	0.1086	0.1013								
	GARCH-IV	0.1044	0.0978	0.0882	0.0736	0.0606	0.0654	0.1603	0.1463	0.1204	0.1058	0.0858								
	GJR	0.1085	0.1120	0.0911	0.0787	0.0541	0.0651	0.1819	0.1769	0.1275	0.0969	0.0974								
	GJR-IV	0.1003	0.995	0.0838	0.0725	0.0557	0.0614	0.1644	0.1536	0.1186	0.0926	0.0894								
	EGARCH	0.0976	0.1002	0.0862	0.0793	0.0507	0.0634	0.1564	0.1498	0.1220	0.0863	0.0897								
	EGARCH-IV	0.0989	0.0956	0.0842	0.0710	0.0546	0.0610	0.1556	0.1437	0.1153	0.0880	0.0873								
	CGARCH	0.1229	0.1236	0.1008	0.0842	0.0582	0.0714	0.2147	0.2050	0.1430	0.1140	0.1044								
	CGARCH-IV	0.1083	0.1061	0.0913	0.0742	0.0590	0.0670	0.1717	0.1633	0.1216	0.1053	0.0949								
	ACGARCH	0.1220	0.1227	0.0994	0.0815	0.0563	0.0681	0.2142	0.2045	0.1391	0.1080	0.0996								
	ACGARCH-IV	0.1090	0.1007	0.0907	0.0742	0.0582	0.0652	0.1733	0.1568	0.1205	0.0989	0.0955								
IV	ARMA	0.1598	0.1350	0.1272	0.1181	0.0847	0.1119	0.2015	0.1736	0.1512	0.1183	0.1334								
	ARMAX	0.1585†	0.1342†	0.1259†	0.1172	0.0846	0.1114	0.1996	0.1735	0.1499	0.1187	0.1322								
	ARIMA	0.1607	0.1357	0.1271	0.1181	0.0846	0.1119	0.2035	0.1752	0.1512	0.1186	0.1341								
	ARIMAX	0.1589	0.1350	0.1259	0.1170†	0.0845†	0.1113†	0.2007	0.1751	0.1502	0.1192	0.1329								
	ARFIMA	0.1604	0.1355	0.1273	0.1183	0.0849	0.1122	0.2013	0.1736	0.1507	0.1181†	0.1332								
	ARFIMAX	0.1590	0.1347	0.1261	0.1175	0.0849	0.1117	0.1991†	0.1733†	0.1491†	0.1185	0.1317†								
	Random walk	0.1601	0.1351	0.1272	0.1178	0.0847	0.1116	0.2023	0.1736	0.1514	0.1191	0.1342								
	ARMA	0.0731	0.0602	0.0643	0.0414	0.0320	0.0297	0.1431	0.1230	0.1221	0.0909	0.0541								
	ARMAX	0.0770	0.0611	0.0652	0.0420	0.0312	0.0304	0.1433	0.1242	0.1238	0.0917	0.0544								
	ARIMA	0.0723	0.0595	0.0574*	0.0386*	0.0294	0.0269*	0.1444	0.1235	0.1226	0.0908	0.0532								
RV	ARIMAX	0.0772	0.0611	0.0587	0.0398	0.0293*	0.0281	0.1452	0.1253	0.1246	0.0898	0.0536								
	ARFIMA	0.0715*	0.0595	0.0590	0.0391	0.0340	0.0285	0.1401	0.1211*	0.1209*	0.0873*	0.0531*								
	ARFIMAX	0.0742	0.0594*	0.0599	0.0400	0.0300	0.0288	0.1396*	0.1215	0.1229	0.0907	0.0534								

Note: The mean absolute forecast error (MAE) and the root mean squared forecast error (RMSE) defined in equations (14) and (15), respectively, of GARCH, IV and RV models are reported. All numbers are multiplied by  $10^3$ . \* denotes the lowest forecast error. † denotes the lowest forecast error among the IV models.

Table 11: Forecast encompassing regression results for the European indices

	GARCH & IV		GARCH & RV		IV & RV		GARCH & IV & RV	
	Model	$adj - R^2$	Model	$adj - R^2$	Model	$adj - R^2$	Model	$adj - R^2$
STOXX	EG-IV & ARX	0.5674	EG-IV & ARX	0.6260	ARFX & ARFX	0.5840	EG-IV & RW & ARX	0.6283
CAC	EG-IV & ARX	0.5894	EG-IV & ARX	0.6383	RW & ARFX	0.5895	EG-IV & ARX & ARX	0.6488
DAX	GJR-IV & ARX	0.6077	GJR-IV & ARX	0.6751	RW & AR	0.6459	GJR-IV & RW & ARX	0.6798
AEX	ACG-IV & ARFX	0.5220	EG-IV & ARX	0.5867	RW & ARX	0.5628	EG-IV & RW & ARX	0.5888
SMI	EG-IV & RW	0.6791	EG-IV & ARX	0.6580	AR & AR	0.5653	EG-IV & RW & ARX	0.6911
FTSE100	GJR-IV & ARX	0.6092	ACG-IV & ARFX	0.6450	RW & ARX	0.5877	ACG & ARFX & ARFX	0.6481

Note: Entries are the encompassing regressions that yield the highest adjusted  $R^2$  values from the encompassing regression described in equation (17).

Table 10: Out-of-sample predictive power of daily volatility forecasts for the European indices

	Models	$adj - R^2$					
		STOXX	CAC	DAX	AEX	SMI	FTSE100
GARCH	GARCH	0.3406	0.3188	0.4524	0.3548	0.4545	0.4308
	GARCH-IV	0.4902	0.4958	0.5122	0.4205	0.4793	0.5530
	GJR	0.4843	0.5143	0.5541	0.4548	0.5999	0.5313
	GJR-IV	0.5274	0.5711	0.6068	0.5015	0.6370	0.6034*
	EGARCH	0.5374	0.5603	0.5769	0.4831	0.6472	0.5390
	EGARCH-IV	0.5626*	0.5890*	0.5951	0.5157	0.6524*	0.6007
	CGARCH	0.2963	0.3188	0.4192	0.3118	0.4193	0.4136
	CGARCH-IV	0.4712	0.4784	0.5209	0.4757	0.5478	0.5536
	ACGARCH	0.2998	0.3226	0.4127	0.3314	0.4564	0.4362
	ACGARCH-IV	0.5196	0.5309	0.5812	0.5147	0.5921	0.6013
GJR IV	ARMA	0.5007	0.4857	0.5201	0.4711	0.4840	0.5276
	ARMAX	0.5133	0.4835	0.5306	0.4804	0.4786	0.5468
	ARIMA	0.4894	0.4753	0.5197	0.4709	0.4830	0.5249
	ARIMAX	0.5107	0.4724	0.5293	0.4784	0.4752	0.5463
	ARFIMA	0.4977	0.4827	0.5196	0.4729	0.4844	0.5272
	ARFIMAX	0.5098	0.4816	0.5296	0.4838	0.4766	0.5444
	RW: $IV_{t-1}$	0.5037	0.4954	0.5171	0.4712	0.4764	0.5413
	RV	ARMA	0.5258	0.5516	0.6330	0.5352	0.5496
ARMAX		0.5238	0.5424	0.6215	0.5277	0.5432	0.5159
ARIMA		0.5181	0.5504	0.6261	0.5365	0.5361	0.5300
ARIMAX		0.5174	0.5377	0.6125	0.5249	0.5267	0.5256
ARFIMA		0.5454	0.5653	0.6371*	0.5456*	0.5476	0.5373
ARFIMAX		0.5471	0.5627	0.6229	0.5344	0.5378	0.5275

Note: Entries are the adjusted  $R^2$  values from the Mincer-Zarnowitz regression described in equation (16). \* denotes the highest adjusted  $R^2$  value.

Table 12: Summary of the 1% VaR failure rates of forecast encompassing regressions

1%							
Model	Ave. failure rate	Sig. Kupiec test	Sig. DQ test	Model	Ave. failure rate	Sig. Kupiec test	Sig. DQ test
1	0.0482	All	All	27	0.0508	All	All
2	0.0400	All	All	28	0.0296	S&P,DJIA,Nasdaq, CAC, AEX, SMI, FTSE	S&P,DJIA,Nasdaq,CAC, AEX, SMI, FTSE
3	0.0458	All	All	29	0.0275	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE
4	0.0428	All	All	30	0.0500	All	All
5	0.0430	All	All	31	0.0300	S&P,DJIA,Nasdaq, CAC, AEX, SMI, FTSE	S&P,DJIA,Nasdaq,CAC, AEX, SMI, FTSE
6	0.0417	All	All	32	0.0504	All	All
7	0.0419	All	All	33	0.0288	S&P,DJIA,Nasdaq, CAC, AEX, SMI, FTSE	S&P,DJIA,Nasdaq,CAC, AEX, SMI, FTSE
8	0.0473	All	All	34	0.0300	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
9	0.0417	All	All	35	0.0464	All	All
10	0.0474	All	All	36	0.0477	All	All
11	0.0433	All	All	37	0.0399	All	All
12	0.0425	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	38	0.0417	All	All
13	0.0422	S&P,DJIA,Nasdaq, AEX, SMI, FTSE	S&P,DJIA,Nasdaq, AEX, SMI, FTSE	39	0.0458	All	All
14	0.0250	All	All	40	0.0433	All	All
15	0.0464	All	All	41	0.0445	All	All
16	0.0233	S&P,Nasdaq, AEX, FTSE	S&P,Nasdaq, AEX, FTSE	42	0.0433	All	All
17	0.0262	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE	43	0.0448	All	All
18	0.0273	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE	44	0.0428	All	All
19	0.0253	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE	45	0.0439	All	All
20	0.0450	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	46	0.0425	All	All
21	0.0296	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE	47	0.0438	All	All
22	0.0464	All	All	48	0.0455	All	All
23	0.0288	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE	49	0.0451	All	All
24	0.0291	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE	50	0.0419	All	All
25	0.0486	S&P,STOXX,CAC, DAX,AEX, SMI, FTSE	S&P,STOXX,CAC, DAX,AEX, SMI, FTSE	51	0.0430	All	All
26	0.0293	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE	52	0.0442	All	All
				53	0.0439	All	All
				54	0.0448	All	All
				55	0.0429	All	All
				56	0.0448	All	All
				57	0.0407	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE
				58	0.0404	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE
				59	0.0220	S&P,DJIA,Nasdaq, STOXX,AEX, SMI	S&P,DJIA,Nasdaq, STOXX,AEX, SMI
				60	0.0191	S&P,DJIA,Nasdaq, STOXX,AEX, SMI	S&P,DJIA,Nasdaq, STOXX,AEX, SMI
				61	0.0462	All	All
				62	0.0239	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE
				63	0.0224	S&P,Nasdaq, AEX	S&P,Nasdaq, AEX
				64	0.0268	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
				65	0.0254	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
				66	0.0255	S&P,DJIA,Nasdaq, AEX, SMI, FTSE	S&P,DJIA,Nasdaq, AEX, SMI, FTSE
				67	0.0252	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
				68	0.0232	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE
				69	0.0214	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE
				70	0.0430	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE
				71	0.0431	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE	S&P,Nasdaq,STOXX, CAC,DAX,AEX, SMI,FTSE
				72	0.0300	S&P,DJIA,Nasdaq, CAC, AEX, SMI, FTSE	S&P,DJIA,Nasdaq, CAC, AEX, SMI, FTSE
				73	0.0271	S&P,Nasdaq, AEX, SMI, FTSE	S&P,Nasdaq, AEX, SMI, FTSE
				74	0.0466	All	All
				75	0.0284	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
				76	0.0265	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
				77	0.0275	S&P,Nasdaq,CAC, AEX, SMI, FTSE	S&P,Nasdaq,CAC, AEX, SMI, FTSE
				78	0.0247	Nasdaq,CAC, AEX, SMI, FTSE	Nasdaq,CAC, AEX, SMI, FTSE

Note: Entries are the 1% average failure rate of the forecasts encompassing regressions. The series for which the Kupiec test for the equality of the empirical failure rate at a specified statistical level and the DQ test for the autocorrelation in VaR violations are significant are listed.

