BONN ECON DISCUSSION PAPERS

Discussion Paper 5/2004

Modeling and Forecasting DAX Index Volatility

by

Zdravetz Lazarov

April 2004

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net

MAIL EXPRESS LOGISTICS FINANCE

Modeling and Forecasting DAX Index Volatility *

Zdravetz Lazarov †

Bonn Graduate School of Economics, University of Bonn

April 15, 2004

Abstract

The recent introduction of the realized variance measure defined as the sum of the squared intra-daily returns stamped on some high frequency basis has spurred the research in the field of volatility modeling and forecasting into new directions. First, the realized variance is a much better estimate of the latent volatility than the sum of the weighted daily squared returns. As such it is better suited for comparing the out-of-sample performances of competing volatility models. Additionally, it can enter as a parameter in these models proving better information than the daily returns commonly used in the standard volatility models. These two innovations have been utilized in several recent papers. We extend this line of research by estimating and comparing a wide class of volatility models for the DAX index futures that use the realized variance or the daily returns. To give a new view of the question whether time series volatility models or implied volatility have better predictive power we estimate a model which incorporates both the historical realized variance and the historical implied volatility. Our results suggest that using realized variance leads to superior performance compared to the previous approaches. Also, the inclusion of the implied volatility produces a slight improvement.

JEL classification: C22; G10

Keywords: Forecasting; High-Frequency Data; Volatility

*The author is grateful to Erik Theissen for providing helpful comments. Deutsche Börse AG generously supplied the data used in this paper.

†Correspondence address: BWL I, Adenauerallee 24-42, D-53113 Bonn, Germany. Tel.: +49 228 73 92 22, E-mail address: zdravetz.lazarov@wiwi.uni-bonn.de .

1 Introduction

Modeling and forecasting volatility has been of significant interest in the field of financial econometrics. Having proper volatility measures and predictions is crucial in the area of risk management (i.e. computing Value-at-Risk) as well as in option pricing and trading. In addition to its practical importance, volatility modeling poses some important theoretical challenges. A general problem that concerns every study of the volatility is that it is inherently unobservable. This rises the need for models that give an exact ex-post estimate of the volatility. Another big area of research is the relation between the option implied volatility and the future volatility. There is largely a disagreement whether historical time series models for volatility or volatility inferred from option prices is better suited for prediction purposes.

The workhorse of the financial econometrics for modeling volatility is still the daily GARCH model. Although many different modifications of the original specification have been introduced, various studies have shown that they provide only a marginal improvement over the standard and mostly used GARCH(1,1) model. GARCH-type of models for the first time provided an ex-post estimate of the volatility in a way that captures successfully the long memory observed in the daily squared returns. In these types of models, the change in the volatility from day to day is inferred using the information in the daily squared returns, which are computed from the closing price of the current day and the closing price of the previous day. Andersen and Bollerslev (1998) criticized this approach, since it ignores the information contained in the intra-daily price movements. They propose a new measure for the daily volatility called the realized variance which is a sum of the intra-daily squared returns computed on a fixed high-frequency basis. In Andersen at al. (2002) it is shown that under very general conditions this measure is an unbiased and efficient estimate of the variance of the daily returns. The authors further claim that the noisiness in the realized variance of the major foreign exchange rates computed at a 5-minutes frequency is so low that it essentially can be used as an error-free measure for both theoretical and practical purposes.

Another important issue concerning volatility is forecasting. Most of the research in that direction compares the performance of the volatility time series models and the implied volatility. Modern literature on that topic started with the seminal paper by Canina and Figlewski (1993). The authors consider the S&P 100 index options market. They regress a measure for the future volatility defined as a weighted sum of the squared returns on the current implied volatility and find little relation between the two variables as indicated by the very low R^2's of the regressions. In addition, when a measure for the historical volatility is added as a regressor then the implied volatility becomes statistically insignificant. In a study using similar methodology Jorion (1995) looks at the market for options in several major currencies. He finds that implied volatility does have relation to the future volatility which is statistically and economically significant even after including the historical volatility in the regression equation. Similar results are reported by Lamourex and Lastrapes (1993) using data on individual equity options. Day and Lewis compare the daily GARCH model and the daily GARCH model with the implied volatility as an exogenous regressor. They find that S&P 100 options implied volatilities do provide information incremental to the historical volatility.

Fleming (1998) and Christensen and Prabhala (1998) criticize the above mentioned studies for using overlapping observations, which induces autocorrelation in the residuals thus leading to inefficient coefficient estimates and possible spurious results. Fleming (1998) constructs a GMM estimator that explicitly accounts for the overlap in the observations while Christensen and Prabhala (1998) use non-overlapping observations in their estimations. Contrary to the previous studies which favor the historical volatility, these authors find that historical volatility does add much incremental information besides that provided by the implied volatility.

Most of these papers use a weighted sum of squared returns as a measure for the ex-post volatility. This measure, however could be quite noisy especially at short horizons. Other studies employ an ex-post estimate of the volatility produced by a parametric model such as GARCH. This approach could induce a serious bias since it favors the model which is used to calculate the estimate for the latent volatility. A few recent papers employ the realized variance measure by Andersen and Bollerslev (1998) instead. For example, Andersen at al. (2002) consider the volatility of the Japanese Yen versus US Dollar and the Deutsche Mark versus US Dollar exchange rates. The authors use an autoregressive fractionally integrated moving average process (ARFIMA) for modeling realized variance. They find that the predictive ability of this model is much better than the predictive ability of the daily GARCH(1,1) model. The authors do not consider the implied volatility. In another paper, Li (2002) compares the implied volatilities and AFRIMA model for the realized variance and finds that historical realized variance does add incremental information besides that provided by the implied volatility. Pong at al.(2002) compare the predictive properties of a short memory (ARMA) model for the realized variance, a long memory (ARFIMA) model for the realized variance and the implied volatility. These authors find that the ARMA model has similar forecasting power as the more complicated ARFIMA model. In addition, the time series models which use historical

data perform better than the implied volatility over short horizons (one day and one week). Over longer horizons (one month and three months) the implied volatility seems to be a better predictor. Martens and Zein (2002) look at the volatility of three separate asset classes: equity, foreign exchange and commodities. They find that in the markets for S&P 500, YEN/USD and Light,Sweet Crude Oil futures time series models for realized variance provide similar and sometimes better forecasts than the implied volatility.
The research so far indicates that there is still largely a confusion about the best way to model and predict volatility. The point of most disagreement is the comparison between the forecasting abilities of the time series models and the implied volatilities. Different studies favor either one of the approaches. However, almost all of them invariably neglect the possibility that the implied volatility incorporates a risk-premium. This may bias any predictions based solely on the implied volatility. In order to account for the volatility risk-premium, a model that incorporates both the historical and the implied volatility is required.
The big discrepancy in the conclusions about the performance of the different volatility models requires new studies which correct for the previous drawbacks. More specifically, the realized variance should be used as a parameter in the time series volatility models. It is much more efficient estimator of the latent volatility than the daily squared returns which enter as parameters in popular volatility models like the daily GARCH model. It also should be used as a benchmark to which the models' performances are compared instead of the widespread practice of using a weighted sum of squared returns or an ex-post GARCH estimate.
Taking all these comments into account we estimate and compare the in- and out-of-sample performances of several volatility models. We focus on the DAX index futures market. This is one the most liquid derivatives markets in the world and the computation of the realized variance measure of Andersen and Bollerslev (1998) is relatively free from market microstructure biases. In addition, there is an active market for DAX options that allows for a robust calculation of the volatility smile, which is of utmost importance for models that utilize the market expectations for the future volatility extracted from option prices.
The first model we consider is the daily GARCH(1,1) model. Additionally, an extension of the daily GARCH(1,1) with the implied volatility is included to correct for the inability of the daily GARCH models to capture abrupt changes in the market conditions. The GARCH-type of models have been initially proposed as a way to account for the high persistence in the squared daily returns. Later, the same specification as the standard GARCH specification has been proposed by Engle and Russell (1998) as a means to model

the durations between transactions, which show a high degree of persistence like the squared daily returns. In a similar spirit, we consider a GARCH model for the realized variance, which to our knowledge has not been done in the literature. A extension of this model with the implied volatilty is also included. Next, we consider an autoregressive fractionally integrated moving average (ARFIMA) model for the realized variance. A couple of recent papers compare the predictions of the ARFIMA model for the realized variance to the predictions obtained from the implied volatility (see Li (2002), Martens and Zein (2002) and Pong at al. (2002)). Instead of comparing both approaches, we investigate the improvement of the forecasts from the ARFIMA model that can be obtained by adding the implied volatility into the model. Finally, we consider an ad-hoc linear regression model, which is defined simply by regressing the realized variance over the corresponding forecasting horizon on the lagged value of the implied volatility.

Our results confirm that the daily GARCH model has significantly worse performance than the models based on the realized variance. Adding the implied volatility changes the output of the model quite a lot. Indeed, the extended daily GARCH model has better forecasting power than the two GARCH models for the realized variance. This is a bit surprising, since the realized variance does not enter as a parameter in this model and at the same time it gives better predictions than the GARCH models based on the realized variance. The GARCH model for the realized variance seems to be well-specified but as we have already mentioned it has worse predictive performance than the daily GARCH model extended with the implied volatility. The inclusion of the implied volatility changes little the output of the model. The ARFIMA model produces better forecasts than the previous models. The extension of this model with the implied volatility improves slightly its performance. The ad-hoc regression model produces unexpectedly good forecasts which are comparable to those of the extended ARFIMA model and at longer forecasting horizons are even better. This is an interesting finding which questions the benefits of using complicated models for forecasting the volatility of the DAX index. More research is needed to see if this holds for other markets.

The wide class of the considered models encompasses the use of the realized variance in time series models, modeling the long term memory of volatility and the simultaneous use of the implied and historical volatilities. This allows to access the importance of these factors for modeling and forecasting purposes. Overall, it seems that in order to produce good forecasts it is important to have a variable which reacts rapidly to changes in market conditions. Our results indicate that the realized and implied volatility almost equally well serve this purpose. Since implied volatility contains a

risk-premium it can not be used directly as a predictor. The extended daily GARCH model which fits the implied volatility to the daily squared returns produces better forecasts than the GARCH model for the realized variance. Also, the ad-hoc regression model produces comparable or better forecasts than the complicated ARFIMA model. Note that for estimating the parameters of this model both the realized and the implied volatility are needed (to account for the volatility risk-premium), but the predictions are made solely on the basis of the implied volatility. This shows that the realized variance contains little incremental information to the implied volatility with regard to the future volatility, but it is possibly needed to infer the volatility risk-premium. On the other hand, adding the implied volatility to the models that utilize realized variance shows little change in their performances.
The rest of the paper is organized as follows. Section 2 contains description of the intra-daily futures data as well as the calculation of the realized variance. Section 3 discusses the algorithm for the calculation of the volatility smile. Section 4 discusses the econometric and economic issues in estimating the relation between the future volatility and the current implied volatility. Section 5 compares the in-sample models' performances while Section 6 compares their out-of-sample predictions over forecasting horizons of lengths between one and ten days. Section 7 summarizes the results and concludes.
Before proceeding further we would like to make the following editorial remark. In the literature 'volatility' traditionally refers to the standard deviation of the returns of a particular asset. However, in a lot of papers the authors implicitly refer to volatility as a measure for the variance instead. In this paper we are concerned with the modeling and predicting the variance of the returns of the DAX index futures. To avoid misunderstanding, we will use the term variance most of the time and the term volatility only in cases when this does not rise confusion.

2 The Futures Data Description and the Computation of the Realized Variance.

We use data provided by Deutsche Bourse on all transactions on the DAX futures and options traded on the electronic trading platform Eurex for the period 4 January 1999 to 31 July 2002 which consists of totally 908 trading days. During that period of time trading started at 8:50 a.m. However, the closing hours have been changed three times. Between 4 January 1999 and 17 September 1999 trading ended at 17:00 p.m., between 18 September 1999 and 1 July 2000 trading ended at 17:30 p.m. and for the rest of the sample the trading ended at 20:00 p.m.

At each point of time there exist three futures contracts with maturities within the cycle March, June, September, December. Contracts expire on the third Friday of the maturity month if that is an exchange trading day, otherwise, on the exchange trading day immediately prior to that Friday.

For the purpose of the computation of the realized variance we use the first to maturity contract, since it is the most liquid and its price follows closely the value of the DAX index. The only exception occurs on the expiration day of the contract when it is traded till 13:00 p.m. In this case we switch to the second-to-maturity futures.

Realized variance measure is the sum of the intra-daily squared returns stamped at some fixed high-frequency. The choice of the frequency is dictated by achieving a balance between measuring the volatility with as little noise as possible on one hand and avoiding market microstructure effects on the other hand. Different studies use returns stamped at intervals of sizes ranging from 5-minutes to 30-minutes. Given the high liquidity of the DAX futures, we use 5-minutes returns. This is also the preferred time interval in the studies for the foreign exchange market.

For each five minute interval we compute the corresponding return as the log-ratio of the value of the last transaction that falls into the interval versus the value of the first transaction that falls into the interval. The realized variance is computed by the formula

$$RV_t^2 = \sum_{t=4}^{l(t)} r_{i,t}^2 + r_t^{*2},$$

where $r_{i,t}$, $i = 1, ..., l(t)$ are the corresponding 5-minutes returns on the $t-$th trading day. We skip the first three 5-minutes intervals(the first 15 minutes of trading) for the following reason. The trading day begins with a pre-opening period when the opening price for the contract is determined. During this period, market participants enter quotes and orders until a time determined by the exchange. After that, a matching of the maximal possible number of quotes and orders is done and an opening price for the contract is determined. This trading mechanism causes sometimes lack of transactions in the very beginning of the trading day which may last up to a quarter of an hour. Overnight volatility is accounted by the term r_t^{*2} which the squared overnight return measured as the log-ratio of the value of the first transaction that falls into the fourth trading interval on the $t+1-$th trading day versus the value of the last transaction on the $t-$th trading day.

After computing the realized variance in the above way we discard from the sample two trading days, namely 11 September 2001 and 20 November 2001.

On both days the variance is unrealistically high and this could distort the econometric analysis. In the former case the realized variance is extremely high and the inclusion of this single observation changes the summary statistics significantly. Since this observation is an outlier caused by non-economic reasons we exclude it to prevent potential bises that could arise when comparing the performance of the different volatility models.
Table 1 presents summary statistics for the realized variance and the squared daily returns measured as the log-ratio of the daily closing prices. Additionally, analogous statistics for the implied variance of the nearest to maturity at-the-money option with time to expiration at least seven days are also presented. As it can be seen the autocorrelation of the realized variance is much higher than that of the squared returns. It also follows a nice monotonically decreasing pattern versus the erratic behavior of the squared returns autocorrelation. The standard deviation of the realized variance is about three times smaller than the standard deviation of the squared returns. This preliminary evidence gives a good indication that the realized variance is a much less noisy measure of the latent variance than the squared returns. Note that the realized variance and the implied variance have similar descriptive statistics.

3 The Options Data Description and the Computation of the Volatility Smile.

We use data on the DAX index options to compute the at-the-money variance and the full volatility smile for near maturity options. This information is used later as a parameter in different volatility models. The data spans all transactions on the DAX options traded on Eurex for the same period as in the case of the futures data. Again we discard two days from the sample, namely 11 September 2001 and 20 November 2001. Trading on the DAX index options ended at the same time as the trading on the DAX futures. However, the trading started ten minutes later at 9:00 a.m.. All options are European style and the underlying is the DAX index. At each point of time there exist contracts with 8 different maturities. For the purpose of this paper we use only the first two nearest to maturity sets of options. They have expiration dates falling on the third Friday (or on the last exchange trading day before if any of these days is a holiday) of the first two successive months. For each trading day we compute the volatility smile implied by nearest to maturity option set if it expires at least seven days from the current day, otherwise we use the second to maturity option set. This is done because very

short to maturity options have implied volatilities that are very sensitive to changes in the price of the underlying and the computation of the volatility smile for such options is rather unstable.
The estimation of the implied volatilities closely follows Hafner and Wallmeier (2000). We partite each trading day into five-minutes intervals. Option prices are matched with the prices of the nearest to maturity futures if the corresponding transactions on the both instruments fall into the same five-minutes interval. We infer the implied index value in the corresponding five-minute interval by the forward-futures parity using the linearly interpolated LIBOR rate that matches the futures maturity. Note that the DAX index is a "total return" index, i.e. it measures the performance of the dividends as well the share returns so the dividend rate is not used in this computation. From the matched option and index prices, the implied volatility is calculated by inverting the Black-Scholes formula. Finally, the volatility smile is computed by fitting a smooth piecewise quadratic function that best matches the computed implied volatilities using the least squares criterion. More precisely, the volatility smile is assumed to have the functional form

$$IV(m) = \alpha_0 + \alpha_1 m + \alpha_2 m^2 + D\left(\beta_0 + \beta_1 m + \beta_2 m^2\right),$$

where:

$$D = \begin{cases} 0, & if\ m \leq 1 \\ 1, & if\ m > 1. \end{cases}$$

We additionally assume that the two segments of the function join at at-the-money point $m = 1$ and that the function $IV(m)$ is differentiable. At the end the shape of the volatility smile reduces to:

$$IV(m) = \alpha_0^* + \alpha_1^* m + \alpha_2^* m^2 + \alpha_3^* D\left(1 + 2 \cdot m + m^2\right).$$

4 The Relation between the Implied Variance and the Latent Variance

A natural candidate for a variable that predicts the future variance is the option implied variance. Intuitively, it could provide a better forecast than the historical time series models since option implied variance is observable and reflects the market expectations for the future variance. However, there are a number of important issues that arise when one uses the implied variance for predictive purposes. The first group of problems concerns the potential biases in the econometric estimation of the relation between the implied variance and the future variance. The second group of problems is related to the extraction of the market expectations for the future variance from option prices. In this section we will give a short discussion of these issues and the ways to deal with them. This will be important later when specifying and evaluating the forecasting performance of the different volatility models.

4.1 Econometric Problems Associated with Estimating the Relation between the Implied and Statistical Variance.

Most of the studies that test the predictive ability of the option implied variances usually employ a regression of the form

$$\sigma^2_{t,T} = \alpha + \beta \cdot \sigma^2_{t,T}\left(bs\right) + \varepsilon_{t,T}, \tag{1}$$

where $\sigma^2_{T,t}\left(bs\right)$ is the implied variance at time t of the at-the-money option with maturity at time T. The dependent variable $\sigma^2_{t,T}$ is an ex-post estimate of the variance over the interval $[t, T]$. The first hypothesis that is tested is whether the implied variance is an unbiased predictor of the future variance, i.e. whether $\alpha = 0$ and $\beta = 1$. This is the so called expectation hypothesis. The second hypothesis that is widely tested is the informational efficiency hypothesis Jorion (1995). It is an encompassing regression which adds a measure for the historical variance $\sigma^2_t\left(hs\right)$ as an additional regressor:

$$\sigma^2_{T,t} = \alpha + \beta \cdot \sigma^2_t\left(bs\right) + \gamma \cdot \sigma^2_t\left(hs\right) + \varepsilon_{t,T}. \tag{2}$$

If the implied variance incorporates all available information at the current moment, it could be expected that the coefficient in front of the historical variance $\sigma^2_t\left(hs\right)$ is not significantly different from zero.

One econometric problem with this regression approach is that most of these tests are performed using overlapping observations, which induces autocorrelation in the residuals $\varepsilon_{t,T}$. While this does not change the point estimates, it could bias the standard errors. Fleming (1998) and Christensen and Prabhala (1998) argue that the use of overlapping data favors the historical variance $\sigma^2_t\left(hs\right)$. These authors point out that this could explain the results of Canina and Figlewski (1993) which show that after including $\sigma^2_t\left(hs\right)$ in the regression (1) the coefficient in front of the implied variance becomes not

significantly different from zero in the new regression (2). Fleming (1998) proposes an instrumental variables solution to this problem and finds that implied variance is an efficient albeit biased estimator of the future variance, which is also informationally efficient. Similar results are reported by Christensen and Prabhala (1998) which use non-overlapping data to avoid this bias. Note that we find that the use of overlapping observations produces similar results as the use of non-overlapping observations when performing regressions of the type (1). A possible explanation for this result is that we employ the realized variance as the ex-post variance estimate $\sigma^2_{t,T}$ instead of a weighted sum of squared returns as Fleming (1998) and Christensen and Prabhala (1998).
Another problem that has received little attention in the literature with the exception of the recent paper by Chernov (2002) is the following. The ex-post estimate of the variance $\sigma^2_{t,T}$ which is usually a weighted sum of the squared daily returns equals the latent variance plus a noise term. If that noise term is correlated with any of the regressors in (1) or (2), this would lead to biased estimates for the coefficients. Chernov (2002) claims that this problem is much more serious than the problem of using overlapping observations. He proposes the use of instrumental variables to alleviate this bias. In his paper he employs a weighted sum of the squared returns as an ex-post measure for the variance $\sigma^2_{t,T}$ and uses its lagged values as instruments. Another way to avoid potential biases of that type is to use a less noisy estimate of the latent variance such as the realized variance. As we mentioned before this has been done in the recent studies by Andersen at al. (2002), Li (2002), Pong at al. (2002) and Martens and Zein (2002). In this paper we also exclusively employ the realized variance as an ex-post variance measure.

4.2 Problems Associated with Inferring the Market Expectations for the Future Variance from the Option Prices.

At-the-money implied variance is widely used as an estimate of the future variance. It is computed by inverting the Black-Scholes formula which assumes that the underlying process follows a geometric Brownian motion with constant volatility. This is a clear inconsistency since if the variance were constant there would not be a need to predict it. In practice the inverting of the Black-Scholes formula is used by the market participants as a transformation of the option prices to the implied volatilities, which are easier to work with. It has been a consensus among traders and academics for many years, without any theoretical justification, that the implied variance(volatility) of the most actively traded at-the-money option is the best market estimate of the future variance(standard deviation). In their seminal paper, Hull and White (1987) for the first time derive a theoretical relationship between the at-the-money variance and the future variance. The authors introduce a stochastic volatility model of the form

$$\frac{\partial S_t}{S_t} = \mu_t \partial t + \sigma_t \partial W_t^1, \tag{3}$$

$$\frac{\partial \sigma_t^2}{\sigma_t^2} = \lambda \partial t + \sigma^2 \partial W_t^2, \ \partial W_t^1 \partial W_t^2 = 0, \tag{4}$$

where the variance of the underlying asset is a stochastic variable and follows a geometric Brownian motion. This model is capable of generating the volatility smile pattern observed in options markets. In that setting Hull and White show that the prices of the plain vanilla options can be expressed as an expectation of the Black-Scholes formula evaluated at the average integrated variance subject to the risk-neutral measure. More specifically

$$HW\left(S_t, K, T, \sigma_t^2\right) = E^Q\left[BS\left(S_t, K, T, V_{t,T}^2\right)\right], \tag{5}$$

where $HW\left(S_t, K, T, \sigma_t^2\right)$ and $BS\left(S_t, K, T, V_{t,T}^2\right)$ are the Hull-White and the Black-Scholes option prices, respectively. Here S_t is the spot price of the underlying, K is the option strike price, T is the time to maturity and σ_t^2 is the spot variance. The term $V_{t,T}^2$ represents the average integrated variance (quadratic variation) over the interval $[t, T]$:

$$V_{t,T}^2 = \frac{1}{T-t} \int_t^T \sigma_s^2 \partial s. \tag{6}$$

Since the Black-Scholes formula is approximately linear in variance for at-the-money options, it follows that for strike prices K near the at-the-money point

$$HW\left(S_t, K, T, \sigma_t^2\right) = E^Q\left[BS\left(S_t, K, T, V_{t,T}^2\right)\right] \approx BS\left(S_t, K, T, E^Q\left[V_{t,T}^2\right]\right), \tag{7}$$

i.e. the implied variance of the at-the-money options is approximately equal to the average future expected variance:

$$\sigma_{t,T}^2(bs) \approx \frac{1}{T-t} E^Q \left[\int_t^T \sigma_s^2 \partial s \right]. \tag{8}$$

If we additionally set the market price of the volatility risk to zero, then the risk-neutral measure in the expectation (8) will coincide with the statistical measure. This immediately implies a relation of the type given by the regression (1).

Note that to justify the regression equation (1) two strong assumptions are needed, namely that the variance follows a geometric Brownian motion, it is uncorrelated with the underlying and the market price for volatility risk is zero. If one of these two conditions is violated, there is no theoretical reason to expect that the implied at-the-money variance is an unbiased predictor of the future variance. In fact, in options markets for each strike price we observe a different value for the implied variance, and in principle each of these values could serve as an estimate for the future variance. In Appendix A we provide a model-free estimate of the future variance implied by the prices of the full spectrum of plain vanilla options having the same expiration date under some relatively weak assumptions. More specifically, at each moment t we derive the mathematical expectation $iv_{t,T}^2$ of the average quadratic variation given by the left hand side of (8) from the prices of the plain vanilla options with maturity at T.

It is instructive to compare the implied at-the-money variance $\sigma_{t,T}^2(bs)$ to the model-free estimation of the future variance $iv_{t,T}^2$. For each day in the sample we calculate the value of $iv_{t,T}^2$ from the nearest to maturity options with time to expiration at least seven days, using the volatility smile whose computation is described in Section 3. When regressing $iv_{t,T}$ on $\sigma_{t,T}^2(bs)$ it turns out that both variables are very closely related to each other:

$$\underset{}{iv_{t,T}} = \underset{(1.12E-06)^{***}}{1.83E-05} + \underset{(0.005)^{***}}{1.013} \cdot \sigma_{t,T}^2(bs) + \varepsilon_t. \qquad \begin{matrix} R^2 = 0.97 \\ N = 906. \end{matrix} \tag{9}$$

The R^2 is very high and the slope coefficient is close to unity. To get a better picture for the economic significance of the bias induced by the intercept in (9), we rescale both variables as the yearly standard deviation (volatility) and run the regression (9) again which yields:

$$\sqrt{iv_t \cdot 365} = \underset{(0.001)^{***}}{0.016} + \underset{(0.005)^{***}}{0.991} \cdot \sqrt{\sigma_t^2\,(bs) \cdot 365} + \varepsilon_t, \qquad \begin{array}{l} R^2 = 0.97 \\ N = 906. \end{array} \tag{10}$$

Again the slope coefficient is quite close to one. The intercept shows that the implied yearly at-the-money volatility underestimates the model-independent estimate $\sqrt{iv_t \cdot 365}$ by 1.6 volatility points on average. Of course this bias may not only be due to the inefficiency of $\sigma_{t,T}^2\,(bs)$ as a predictor of the future variance, but also due to inaccuracies in the computation of $iv_{t,T}^2$. The calculation of $iv_{t,T}^2$ is done using the prices of both in- and out-of-the money options, which have higher bid-ask spreads than the at-the-money option. This makes the room for errors in the calculation of $iv_{t,T}^2$ significantly higher than in the calculation of $\sigma_{t,T}^2\,(bs)$. Nevertheless both variables are very closely related to each other up to a constant term. Since the models we consider are invariant to a linear transformation of the implied variance, in what follows we use only the at-the-money implied variance. It has already been widely used in the literature and also allows for more straightforward and error-free calculation.

Another problem which arises when comparing the latent statistical variance to the implied variance is that every estimation for the future variance inferred from option prices incorporates a risk premium. Unless the market price for the variance risk is zero, any regression of the from (1) will be biased i.e it will give estimates for which $\alpha \neq 0$ and $\beta \neq 1$. This possibility is almost always neglected in the literature. Unfortunately, the only way to gauge the impact of the variance risk premium is to specify a parametric model. To show how a non-zero risk premium can affect the regression (1), we will give a short example using the well-known Heston model. The price dynamics of the underlying instrument and its variance under the statistical measure P in the Heston model are given by:

$$\frac{\partial S_t}{S_t} = \mu_t \partial t + \sigma_t \partial W_t^1, \tag{11}$$

$$\partial \sigma_t^2 = \left(\theta - k\sigma_t^2\right)\partial t + \sigma\sqrt{\sigma_t^2}\partial W_t^2,\ \partial W_t^1 \partial W_t^2 = \rho. \tag{12}$$

The market price of the variance risk is proportional to the spot volatility $\Lambda_t = \varphi \cdot \sqrt{\sigma_t^2}$. Under the risk-neutral measure the dynamics of the variance is given by:

$$\partial \sigma_t^2 = \left(\theta - (k + \varphi)\,\sigma_t^2\right)\partial t + \sigma\sqrt{\sigma_t^2}\partial W_t^2. \tag{13}$$

Using well-known facts for the mean-reverting processes (12) and (13), the expectations $E^P\left[\int_t^T \sigma_s^2 \partial s\right]$ and $E^Q\left[\int_t^T \sigma_s^2 \partial s\right]$ can be readily calculated:

$$E^P\left[\int_t^T \sigma_s^2 \partial s\right] = \alpha_{t,T} + \beta_{t,T} E^Q\left[\int_t^T \sigma_s^2 \partial s\right] + \gamma_{t,T}\sigma_t^2. \qquad (14)$$

The coefficients $\alpha_{t,T}$,$\beta_{t,T}$ and $\gamma_{t,T}$ depend only on t and T and the parameters in variance specifications (12) and (13). The last expression implies a regression of the form (1) with $\alpha \neq 0$ and $\beta \neq 1$ plus an additional regressor for the spot variance. From this example, it is clear that tests of the predictive ability of the implied variance using regressions of the type (1) could be misleading. The existence of a variance risk-premium could make the relation between the implied and statistical variance more complicated than that implied by the regression (1) with the conditions $\alpha = 0$ and $\beta = 1$. Chernov (2002) finds that by estimating a regression of the form (1) with the spot variance as an additional regressor as postulated by the relation (14) and taking into account the possibility for non-zero intercept and slope coefficient different from unity, the expectation hypothesis for the implied variance can not be rejected. Motivated by this result, in Section 6 we employ a regression model suggested by the relation (14) for predictive purposes. However, in contrast to the findings obtained by Chernov (2002), the coefficient in front of the spot variance turns out to be insignificant.

5 In-Sample Modeling of the DAX Volatility

Before proceeding to the issue of forecasting, it is important to have a good picture of the ex-post estimates of the variance produced by different volatility models. This will serve as an additional guide in evaluating their out-of-sample forecasting abilities.

In this section we consider the in-sample performances of several models. It turns out that the estimates produced by the standard daily GARCH model differ significantly from the models which use the realized variance or the implied variance. This difference is due to use of the daily squared returns in the daily GARCH model which contain a lot of noise and are not able to capture abrupt changes in the market volatility. This is also confirmed by the worse out-of-sample performance of this model reported in the next section.

5.1 The Daily GARCH Model and the Daily GARCH Model Enhanced with the Implied Variance

The daily GARCH model is still the most widely used volatility model. It allows to infer the latent variance from the squared daily returns. The standard GARCH(1,1) model specifies the variance of the daily returns σ_t as a predictable process determined by a linear combination of the previous squared returns r_{t-1}^2 and the previous conditional variances σ_{t-1}^2:

$$r_t = \varepsilon_t \cdot \sigma_t, \quad \varepsilon_t \text{ are i.i.d,} \tag{15}$$

$$\sigma_t^2 = \alpha + \beta \cdot r_{t-1}^2 + \gamma \cdot \sigma_{t-1}^2. \tag{16}$$

This specification is able to account for the long term autocorrelation of the squared daily returns, which is captured by the lagged conditional variance value σ_{t-1}^2 in (16). The current changes in the variance are accounted by the lagged value of the squared daily return r_{t-1}^2.

The use only of daily data has drawn some criticism since it contains a lot of noise and the changes in the market conditions are inefficiently incorporated. One way to alleviate this problem is to include exogenous variables in (16). A good candidate for such variable is the at-the-money implied variance. As we have already commented, it is widely thought to reflect the market expectations for the future variance by aggregating all currently available information. Additionally, our results in the previous section show that it is very closely related to the model independent prediction $iv_{t,T}^2$ for the future variance. The idea of using the implied variance is utilized by Day and Lewis (1992). The authors look at the information content of the implied variance relative to the information contained in the squared daily returns by including it as an exogenous variable in the GARCH and EGARCH models. Overall, they find that including the at-the-money implied variance improves the models' performances.

We also carry out similar analysis to that of Day and Lewis (1992). In Table 2 are shown the estimation results for the GARCH(1,1) model as well as for the GARCH(1,1) model enhanced with the implied variance $\sigma_{t-1}^2\,(bs)$ of the nearest to maturity at-the-money option with time to expiration at least seven days. Apparently, the inclusion of $\sigma_{t-1}^2\,(bs)$ as an exogenous variable in (16) appears to capture the long-term memory property of the variance, since the coefficients in front of the conditional lagged variance σ_{t-1}^2 are not significantly different from zero. The coefficient in front of the lagged values of the squared daily return is significant and negative. This possibly indicates

that the implied variance contains a risk-premium which is proportional to the level of the latent variance.
To compare the output gv_t^2 of the daily GARCH model and the output $gv\,(bs)_t^2$ of the extended daily GARCH model we regress gv_t^2 on $gv\,(bs)_t^2$:

$$\begin{array}{lcccll} gv_t = & \underset{(7.76E-06)^{***}}{7.71E-05} & + & \underset{(0.035)^{***}}{0.754} & \cdot gv\,(bs)_t^2 + \varepsilon_t, & \begin{array}{l} R^2 = 0.50 \\ N = 905. \end{array} \end{array} \quad (17)$$

The slope coefficient is significantly below one and the intercept is significantly different from zero. The R^2 of the regression is also relatively low. Thus, adding the implied variance as an exogenous regressor in the daily GARCH specification changes the estimation output significantly.

5.2 Realized Variance and the GARCH-RV Model

In this section we propose an autoregressive conditional model for the realized variance which is analogous to the standard daily GARCH model and the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998). This model allows for a convenient and parsimonious way to capture the long-term memory observed in the realized variance series.
The model starts with specifying the joint density of the realized variances RV_1^2,....,RV_n^2 as a function $p\,(x_1, ..., x_n; \vartheta)$ depending on a set of parameters ϑ . This density function can be written as a product of the successive conditional densities:

$$p\,(x_1, ..., x_n; \theta) = \prod_{t=1}^{n} p\,(x_t | x_{t-1} ..., x_1; \vartheta)\,. \quad (18)$$

The idea is, as in Engle and Russell (1998) to specify the functional form of the conditional densities directly such that the long path-dependency of the realized variance is captured. More specifically, it is assumed that there exist a function h_t which depends only on the information available up to the moment $t-1$ such that the scaled realized variances $\frac{RV_t^2}{h_t}$ are identically and independently distributed. In that way the value of h_t summarizes the information contained in the previous movements of the realized variance. The function h_t is set to the conditional expectation of RV_t^2 at the moment $t-1$:

$$h_t = E_{t-1}\left[RV_t^2\right], \quad (19)$$

with the functional form

$$h_t = \alpha + \beta \cdot RV_{t-1}^2 + \gamma \cdot h_{t-1}. \tag{20}$$

We call this specification the GARCH-RV(1,1) model. It is analogous to the GARCH(1,1) model and the ACD(1,1) model of Engle and Russell (1998). All three models have the same likelihood function. One can add more lagged values in (20) of both the realized variance and the conditional realized variance but the estimation of this extended specification shows that there is no difference in the performance as in the case of the daily GARCH model. The variable h_t can be interpreted as the one step ahead prediction of the realized variance. Another alternative, if we assume that the realized variance is not an error-free estimate of the latent variance, is to assume that h_t is the ex-post value of the latent variance at time t as in the standard GARCH specification.

Table 2 presents the estimation results for the GARCH-RV model and Table 3 shows some summary statistics for the scaled realized variances $\frac{RV_t^2}{h_t}$. The null hypothesis that the mean value of $\frac{RV_t^2}{h_t}$ is one can not be rejected. Also, the lagged autocorrelations of the scaled variances are not significantly different from zero. Overall, it seems that the GARCH-RV is successful in capturing the long-memory property of the realized variances.

It is interesting to compare the conditional expected values h_t and the realized variances RV_t^2 by running the following regression:

$$\begin{array}{llllll} RV_t^2 = & 1.44E-05 & + & 0.966 & \cdot h_t + \varepsilon_t, & R^2 = 0.40 \\ & (1.91E-05) & & (0.088)^{***} & & N = 906. \end{array} \tag{21}$$

The intercept is not significantly different from zero and the slope of the regression is close to unity, which confirms that the GARCH-RV model is correctly specified. Although the expected value of the realized variance and its conditional expectation are close to each other, the difference between their values is economically significant as indicated by the 40 percent value of R^2. Additionally, the expected conditional realized variance is much more persistent and has smaller standard deviations than the realized variance. In that respect conditional realized variance is more similar to the daily GARCH(1,1) variance and the implied variance.

We also estimate the GARCH-RV model extended with the implied variance. The estimation results in Table 2 show that although the coefficient in front of the lagged implied variance is significant the log-likelihood function increases negligible compared to the GARCH-RV model. Thus, it seems that

the realized variance is successful in capturing similar effects that are captured by the implied variance. This results is in contrast with the significant increase of the likelihood function when adding the implied variance as a exogenous regressor in the daily GARCH model.
Finally, it is interesting to compare the in-sample outputs of the models considered in this section. Table 4 presents the regression estimations derived by regressing the estimates of the latent variance of each of the models on the variance estimates of the remaining models. As it can be seen, regressions in which the daily GARCH variance is either a dependent variable or a regressor have the lowest R^2's and the slope coefficients are markedly lower than one. The GARCH-RV and the extended GARCH-RV models produce similar outputs as indicated by the high R^2's and the slope coefficients close to one in the corresponding regressions. Table 4 also shows that the extended daily GARCH model have closer behavior to the two GARCH-RV models than to the daily GARCH model.
The difference in the in-sample performance of the daily GARCH model to the performances of the other models, as well as its lowest likelihood value suggests the importance of using the realized variance and the implied variance. The results in the next section confirm this intuition.

6 Evaluating the Forecasting Performance

One of the major challenges in testing the predictive ability of the volatility models is that the forecasted variable is unobservable. In practice the predicted value of the variance is compared to an ex-post estimate not to the exact value. Previous studies usually use an ex-post model-dependent estimate of the latent variance as a benchmark for very short forecasting horizons such as one day. This practice clearly could induce a bias since it is possible that it favors the out-of-sample performance of the model which is used to produce the ex-post variance values. For longer forecasting horizons a weighted sum of the squared daily returns is mostly used. While this gives a model-free benchmark value, it could be a quite noisy measure. To avoid these drawbacks, we employ the realized variance as a benchmark to compare the out-of-sample performances of the different models. It has the advantage that it is model-free and contains little noise.
First, we consider a one-day forecasting horizon. The models we compare include the daily GARCH model, the daily GARCH model enhanced with the implied variance, the GARCH-RV model, the GARCH-RV model enhanced with the implied variance. Additionally, we consider two autoregressive fractionally integrated moving average (ARFIMA) models. Fractional

integration approach has been shown to be useful in modelling the long-term memory of variety of economic time series, including the realized variance. We estimate and produce forecasts for the ARFIMA model for the realized variance as well as an extended version of the model with the implied variance. Finally, an ad-hoc regression for the implied variance is estimated and used for out-of-sample predictions.

Next, we consider forecasting horizons with lengths between two and ten days. The models we compare are as before with the exception of the extended daily GARCH and the extended GARCH-RV models. The reason is that more than one day predictions in these two models require specifying the dynamics and forecasting of the implied variance. On the other hand, multi-step predictions in the extended ARFIMA model are easier to implement as explained further in the paper.

6.1 One-day Forecasting Horizon

The approach used to produce out-of-sample predictions is to estimate the corresponding model up to day t and on that basis to form an out-of-sample prediction for the day $t+1$. To obtain precise estimation of the models' parameters we make predictions only for the days $t+1 \geq 401$, which guarantees at least 400 in-sample observations. The quality of the forecasts is measured by the Heteroscedascity-consistent Mean Square Error defined as:

$$HRMSE = \sqrt{\frac{1}{n-k+1}\sum_{t=k}^{n}\left(1-\frac{Forecast_{t-1,t}}{RV_t}\right)^2}, \tag{22}$$

where $Forecast_{t-1,t}$ is the variance forecast on day $t-1$ for the day t and $[k,n]$ is the interval used for predictions. In our case $k=401$ and $n=906$. The HMSRE measure is a modification of the standard RMSE measure with the additional advantage that it provides robustness to serial correlation in the residuals. It has been used in the context of evaluating volatility models by Andersen and Lange (1999) and Martin and Zein (2002). Another popular approach to evaluate the forecasting performance is to regress the realized variance on its predicted value

$$RV_t = \alpha + \beta \cdot Forecast_{t-1,t} + \varepsilon_t \tag{23}$$

and then to compare the R^2's produced by the forecasts of the different models. One should be careful however, since the high persistence of the realized

variance and the forecasted values $Forecast_{t-1,t}$ could result in a spurious explanatory power. We use this regression approach here merely as a robustness check of the model evaluation given by the criterion (22). It turns out that both criteria (22) and (23) give similar results.
We have already considered the daily GARCH and the GARCH-RV models as well as their extended versions in Section 5.1. Motivated by the discussion in Section 4.2, an ad-hoc linear regression is also used for predictions. Forecasts are made in the following way. The estimates of the coefficients $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ of the regression

$$RV_k^2 = \bar{\alpha} + \bar{\beta} \cdot \sigma_{k-1}^2 \left(bs\right) + \bar{\gamma} \cdot RV_{k-1}^2 + \varepsilon_k \tag{24}$$

are obtained using the observations $k = 2, ..., t$. Here RV_{k-1}^2 serves as a proxy for the spot variance on day $k-1$. The one day prediction of RV_{t+1}^2 on day t is given by $\bar{\alpha} + \bar{\beta} \cdot \sigma_t^2 \left(bs\right) + \bar{\gamma} \cdot RV_t^2$. It turns out that in all estimations for $t + 1 = 401, ..., 906$, the coefficient $\bar{\gamma}$ is never significant and by that reason the predictive regression (24) is estimated without the regressor RV_{k-1}^2.
The last model we consider is the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model for the realized variance. A process y_t is said to be an $ARFIMA(d, p, q)$ if the fractionally integrated process $\left(1 - L\right)^d y_t$, $0 \leq d \leq 1$ is a stationary $ARMA\left(p, q\right)$ process. The fractionally integrating operator $\left(1 - L\right)^d$ is defined as:

$$\left(1 - L\right)^d = 1 + d_1 L^1 + d_2 L^2 + ..., \tag{25}$$

where the coefficients d_k are given by the formula:

$$d_k = \frac{d\left(d-1\right)...\left(d-k+1\right)}{k!}. \tag{26}$$

The decomposition (25) is an exact analogue of the binomial decomposition $\left(1 - x\right)^d = 1 + d_1 x^1 + d_2 x^2 + ..., x \in R$. Intuitively, the parameter d captures the long-term memory of the process in a way that the operator $\left(1 - L\right)^d$ reduces the long-path dependency of the series and the remaining short term path-dependency in the transformed series $\left(1 - L\right)^d y_t$ is captured by the ARMA specification. If $d = 0$ then y_t is an ARMA process and if $d = 1$ then y_t is an ARIMA process.
As in the previous applications of the ARFIMA model to volatility modeling (see Andersen at al. (2002), Li (2002), Pong at al. (2002) and Martens and Zein (2002)) we employ the log values of the realized variance $y_t = \log rv_t^2$

since they show more regular behavior than the realized variances. Indeed, the unconditional distribution of the log realized variances is very close to the normal one. We estimate the degree of fractional integration of the log realized variance using the Robinson (1995) non-parametric approach. The estimations give values which are close to 0.40. This is in line with the estimations produced by Andersen at al. (2002) for the FX market. The result is stable across sub-samples. In what follows we set the degree of fractional integration of the log realized variance to 0.40 in all of the further estimations. It seems that most of the observed autocorrelation of log realized variance $\log(rv_t)$ is caused by the long memory of the process. After performing the one parameter transformation $(1-L)^d \log(rv_t)$, which accounts for the long-memory of the process, the resulting $ARMA(p,q)$ model, that reflect the short memory of the realized variance, has an R^2 of only four percent. Similar results are obtained by Andersen at al. (2002) for the major foreign exchange rates.

The AFRIMA model as specified above uses only the historical log realized variances as inputs. One may wish to include the historical implied variances as additional parameters. The easiest way to do that is to assume that the fractionally integrated process $(1-L)^d \log(rv_t)$ is an $ARMA(p,q)$ with implied variances as additional exogenous regressors. However, this assumption is inconsistent with the methods for estimating the degree of fractional integration. Indeed, it turns out that the inclusion of the lagged values of the implied variances into the ARMA specification for the fractionally integrated log realized variances $(1-L)^d \log(rv_t)$ induces autocorrelation into the new standardized residuals[1].

To make the things self-consistent and correctly specified we use a two-step estimation procedure. The idea is first to estimate an ARFIMA model for the log realized variances and then check if there is some predictability in the standardized residuals induced by the lagged values of the fractionally integrated log implied variances. For a formal description of this approach, see Appendix A.

Table 5 shows the forecasting performance of the models. Both HRMSE and R^2 criteria produce the same ranking. The only exception is the linear regression model. It is ranked first according to the R^2 criterion and third according to the HRMSE criterion. This discrepancy could be explained by the fact that this model is simply a regression fitted to the historical data so it is not surprising that it produces good forecasting results based on the R^2 criterion. As we mentioned, before the HRMSE criterion is used to rank the models and R^2 criterion serves merely as a robustness check.

[1] See Appendix A for more details on that point

The best performing model is the ARFIMA model enhanced with the implied variance. This suggests that both the historical and the implied variances are important in predicting the future variance. The next best performing model is the ARFIMA model. This confirms that the fractional integration approach captures well the dynamics of the realized variance. Very close performance in terms of HRMSE error has the ad-hoc linear regression. While it is much easier to estimate than the ARFIMA model, it uses the implied variance as an additional parameter. Surprisingly, the next best performing model is the daily GARCH model extended with the implied variance. Note that the realized variance does not enter as a parameter in this model and at the same time it is capable of better predicting the realized variance than the GARCH-RV and the extended GARCH-RV models. The superior performance of the extended GARCH-RV model over the GARCH-RV model suggests again the importance of the implied variance. Finally, the worst performing model according to both HRMSE and R^2 criteria is the daily GARCH model. Its predictive power is significantly lower than that of the other models.

A closer look at the regression estimates of the realized variance on its predicted values in Table 5 could provide us with some additional insights into the models' performances. First, note that all intercept coefficients are not significantly different from zero. Second, the slope coefficients are generally close to unity, with the exception of the daily GARCH model and both ARFIMA models. In the case of the daily GARCH model, this effect is due to its general inability to model the variance in- and out-of-sample which has been already discussed. The high slope coefficients in the regressions corresponding to the ARFIMA models can be explained by the fact that these two models forecast the log values of the realized variance, which are later transformed to predictions for the realized variance by taking an exponent. This leads to a bias of the following form. Suppose that $\widetilde{rv}_{t,t+1}$ is the unbiased one-day ahead prediction of the log realized variance $\log(rv_{t+1})$ produced by one of the two ARFIMA model, i.e. $\log(rv_{t+1}) = \widetilde{rv}_{t,t+1} + \varepsilon_{t+1}$, where $E[\varepsilon_{t+1}|I_t] = 0$. Here I_t is the information available up to the day t. The value that is used as a one-day ahead forecast is $\exp(\widetilde{rv}_{t,t+1})$. On the other hand, the conditional expectation of rv_{t+1} is given by $E[rv_{t+1}|I_t] = \exp(\widetilde{rv}_{t,t+1}) \cdot E[\exp(\varepsilon_{t+1})|I_t]$. The two values differ from each other by a factor of $E[\exp(\varepsilon_{t+1})|I_t]$. Computing the correct forecast requires the calculation of $E[\exp(\varepsilon_{t+1})|I_t]$. If we assume that the residuals ε_{t+1} are conditionally normally distributed with variance σ^2_{t+1} then $E[\exp(\varepsilon_{t+1})|I_t] = \exp\left(\frac{1}{2}\sigma^2_{t+1}\right)$. However, we find that the modeling of the conditional variance of the residuals in our ARFIMA setting to be extremely challenging. For that reason we take the exponent of the prediction of the log realized variance and use it as a forecast for the realized variance. To our knowledge, none of the papers that uses ARFIMA type of modeling for the realized variance mentions explicitly how the predictions of the realized variance are formed from the predictions of the log realized variance. Our guess is that this is done simply by taking the exponent as the approach pursued here.

6.2 Two to Ten Days Forecasting Horizon

In this section we compare the models' predictions for horizons with lengths between two and ten days. For each forecasting horizon with length $l = 2, ..., 10$, we estimate the corresponding model up to day t and on that basis form a prediction of the realized variance over the next l days $RV_{t+1}^2 + ... + RV_{t+l}^2$, where $t = 400, ..., 906 - l$. As before, the different forecasts are compared using the HRMSE criterion (22). Additionally, the R^2 values from regressions of the type (23) are computed. Since previous authors have noted that using overlapping observations could lead to biased results, first we run the regressions (23) over l sets of non-overlapping observations for each forecasting horizon with length $l = 2, ..., 10$. However, it turns out that the coefficient estimates and R^2 values do not differ much from those computed using single regressions that span all available observations. For that reason we report the estimates obtained by employing regressions of the latter type.

The models we consider are as before with the exception of the extended daily GARCH model and the extended GARCH-RV model. As we have already mentioned, predictions for more then one day ahead in these models requires specifying the dynamics of the implied variance. One the other hand multi-step forecasts can be readily introduced in the extended ARFIMA framework since the implied variance is an white noise when it is integrated at a level of fractional integration $d = 0.9$. For a detailed description of the forecasting in the extended ARFIMA model see Appendix B.

Table 6 presents the models' performances in terms of the HRMSE and R^2 criteria. First, notice that the daily GARCH model has significantly less predictive power than all of the other models. It has the worst performance over all forecasting horizons according to both criteria. Although the GARCH-RV model performs significantly better than the daily GARCH model, its predictive abilities are also worse than that of the two ARFIMA models and the linear regression model. The two ARFIMA models have similar performances. However, the extended ARFIMA model always produces slightly better or similar predictions according to both HRMSE and R^2 criteria. Its advantage over the ARFIMA model decreases with the length of the fore-

casting horizon. The linear regression model shows similar although slightly worse performance compared to the extended ARFIMA model for forecasting horizons with length less than six days. For longer forecasting horizons it becomes the best performing model with slightly better predictions than the two ARFIMA models.
A look at the Table 7 reveals that regressing the realized variances on its predicted values gives biased results for all models with the exception of the GARCH-RV model. The daily GARCH model produces slope coefficients significantly less then unity. In addition, the intercept coefficients are also significantly different from zero for forecasting horizons greater than four days. On the other hand, the GARCH-RV model has the smallest bias with slope coefficients close to unity and intercept coefficients, which are always insignificantly different from zero. As in the case of the one-day forecasting horizon, the two ARFIMA models have slope coefficients rather bigger than unity and often have intercept coefficients significantly different from zero. We have already discussed that this is due to the use of the log value of the realized variance in these models. The linear regression model has always insignificant intercept coefficients and relatively close to unity slope coefficients which increase with the forecasting horizon.

7 Summary and Conclusion

In this paper we analyze the in- and out-of-sample performances of several volatility models while accounting for some of the drawbacks in the previous research. We find that the daily GARCH model has significantly worse performance than the other models. In particular, it produces weaker forecasts than the models using implied variance. This is in accord with the findings of Christensen and Prabhala (1998), Fleming (1998), Jorion (1995), Day and Lewis (1992) and Martens and Zein (2002), who consider several different markets including the S&P 100 and S&P 500 index markets, the FX market, and the Light,Sweet Crude Oil futures market. The daily GARCH model also has a worse performance than the ARFIMA model for the realized variance which is consistent with Andersen et al. (2002) and Martens and Zein (2002). Adding the implied variance as an exogenous regressor dramatically improves the predictive ability of the daily GARCH model. We also consider GARCH-type of modeling for the realized variance in the spirit of Engle and Russell (1998) which to our knowledge has not been done in the literature. Econometric tests show that the GARCH-RV model for the realized variance is correctly specified and produces unbiased results. Adding the implied variance as an exogenous regressor in this model changes its output little in contrast to the daily GARCH model. This again confirms that realized and implied variance are superior to daily squared returns for modeling volatility. Although the GARCH-RV model has better predictive power than the daily GARCH model, it performs worse than the ARFIMA model and the ad-hoc linear regression model.

Our analysis shows that fractionally integrating the log realized variance captures its path dependent properties surprisingly well since the ARMA model for the fractionally integrated log realized variance has a very low R^2. Similar findings are obtained by Andersen et al. (2002). Additionally, it seems that the ARFIMA model is also very suitable for modeling the log implied variance, which integrates to white noise for a degree of fractional integration of 0.90. Although, as in the studies of Andersen et al. (2002), Li (2002) and Martens and Zein (2002), we find that the ARFIMA model has a good forecasting ability, it is nevertheless slightly improved by using the implied variance as an additional parameter.

Our paper leads to two conclusions which enrich the current literature on modeling and forecasting volatility. First, it seems that the implied and the realized variance contain almost the same information with regard to the future variance. On one hand, adding the implied variance to models that employ the realized variance such as the GARCH-RV model and the ARFIMA model changes their performances either insignificantly or very little. On the other hand, models based on the implied variance perform almost as good or slightly better than the models based on the realized variance. Remember that the daily GARCH model extended with the implied variance performs better than the GARCH-RV model, and the realized variance does not enter as a parameter in this model. The ad-hoc linear model performs comparably to the ARFIMA model. Note that although this regression model is estimated using both the implied and the realized variance, the predictions are made using only the implied variance. The estimation of this model requires the use of the realized variance in order to account for the volatility risk-premium. This results may help reconcile the current split in the literature concerning the predictive properties of the volatility time series models and the implied variance. It is possible that sometimes the implied variance forecasts are worse than the historical time series forecasts, because of the volatility premium. Therefore, to make predictions on the basis of the implied variance, it should be fitted to the historical data in order to isolate the impact of the volatility risk-premium.

Second, we show that the ad-hoc linear regression model has a surprisingly good performance. This model is very easy to estimate, given that estimates for the realized and implied variances are available. It produces forecasts which are comparable to the ARFIMA model extended with the implied variance for forecasting horizons up to six days. For longer forecasting horizons it gives the best forecasts. Given the simplicity of this model, it may well be the preferred choice for performing specific practical tasks, such as computing VaR. Future research is needed to see if these two conclusions hold for other markets.

References

[1] Andersen, T.B., and T. Bollerslev. "Answering the skeptics: Yes standard volatility models do provide accurate forecasts." International Economic Review, 39 (1998), 885-900.

[2] Andersen, T.B., T. Bollerslev, F.X. Diebold, and P. Labys. "Modeling and forecasting realized volatility", Econometrica 71 (2003), 579-625.

[3] Andersen, T.G., T. Bollerslev, and S. Lange (1999), "Forecasting financial market volatility: sampling frequency vis-a-vis forecast horizon." Journal of Empirical Finance, 6 (1999), 457-477.

[4] Breeden, D. and R. Litzenberger. "Prices of state contingent claims implicit in option prices.", Journal of Business, 51 (1978), 621-651

[5] Canina, L., and S. Figlewski. "The information content of implied volatility." Review of Financial Studies, 6 (1993), 659-681.

[6] Chernov M. "On the role of volatility risk premium in implied volatilities based forecasting regressions." Working paper, Columbia University (2002)

[7] Christensen, B.J., and N.R. Prabhala. "The relation between implied and realized volatility." Journal of Financial Economics, 50 (1998), 125-150

[8] Day T., and C. Lewis. "Stock market volatility and the information content of stock index options." Journal of Econometrics, 52 (1992), 267-287.

[9] Engle, R., and J. Russell. "Autoregressive conditional duration: a new model for irregularly spaced data." Econometrica, 66 (1998), 1127-1162.

[10] Fleming, J. "The quality of market volatility forecasts implied by S&P 100 index option prices." Journal of Empirical Finance, 5 (1998), 317-345

[11] Green, R.C. and R.A.Jarrow. "Spanning and completeness in markets with contingent claims." Journal of Economic Theory, 41 (1987), 202-210

[12] Hafner,R. and M.Wallmeier. "The dynamics of DAX implied volatilities." Proceedings of the 13th Annual CBOT Research Symposium. (2000)

[13] Hull J.C. and A.White. "The pricing of options on assets with stochastic volatilities" Journal of Finance, 42 (1987), 381-400

[14] Jorion P. "Predicting volatility in the foreign exchange market." Journal of Finance, 50 (1995), 507-528

[15] Li K. "Long memory versus option-implied volatility predictions." Journal of Derivatives, 9 (2002), 9-25

[16] Martens, M. and J.Zein. "Predicting financial volatility: high-frequency time-series forecasts versus implied volatility." Working paper, Erasmus University (2002)

[17] Pong. S., M.B. Shackleton, S.J. Taylor, and X. Xu. "Forecasting Sterling/Dollar volatility: implied volatility versus long memory intraday models." Working paper Lancaster University (2002)

[18] Robinson, P.M. "Long-periodogram regression of time series with long-range dependence." The Annals of Statistics, 23 (1995), 1048-1072

[19] Ross, S.A. "Options and Efficiency." Quarterly Journal of Economics, 90 (1976), 75-89.

A Appendix - Model-Free Estimation of the Future Variance

We will derive a model-independent estimate of the future variance from the option prices under some relatively weak assumptions. This estimate represents the mathematical expectation of the quadratic variation of the variance under the risk-neutral measure. To fix ideas, let $F_s, t \leq s \leq T$, be the price of the futures contract with maturity at time T . We assume that the process for the futures price is a continuous semi-martingale and a full spectrum of plain vanilla options with maturity at time T exists. Additionally, the interest rate is set to a constant equal to r. Let f be an arbitrary twice differentiable function. From the Ito's lemma it follows:

$$f(F_T) - f(F_t) = \int_t^T f'(F_s)\,\partial F_s + \int_t^T \frac{F_s^2}{2} f''(F_s)\,\sigma_s^2 \partial s. \tag{27}$$

Under the risk-neutral measure Q the futures price process F_s is martingale, so it follows:

$$E^Q\left[f(F_T) - f(F_t)\right] = E^Q\left[\int_t^T f'(F_s)\,\partial F_s\right] + E^Q\left[\int_t^T \frac{F_s^2}{2} f''(F_s)\,\sigma_s^2 \partial s\right]$$

$$= E^Q\left[\int_t^T \frac{F_s^2}{2} f''(F_s)\,\sigma_s^2 \partial s\right]. \tag{28}$$

Neuberger (1990) first noticed that by setting $f(F_s) = 2 \cdot \log(F_s)$ then $\frac{F_s^2}{2} f''(F_s) \equiv 1$ and expectation in the right hand side of (28) equals the expected future variance under the risk neutral measure. More specifically:

$$E^Q\left[\int_t^T \sigma_s^2 \partial s\right] = 2 \cdot E^Q\left[\log(F_T)\right] - 2 \cdot \log(F_t). \tag{29}$$

Applying the forward-futures parity $F_t = S_t e^{r(T-t)}$[2] expression (29) can be written as:

[2]Remember that the dividends are included in the calculation of the DAX index

$$E^{Q}\left[\int_{t}^{T}\sigma_{s}^{2}\partial s\right]=2\cdot e^{r(T-t)}\frac{E^{Q}\left[\log\left(F_{T}\right)\right]}{e^{r(T-t)}}-2\cdot\log\left(S_{t}\right)-2\cdot r\left(T-t\right). \quad (30)$$

Since $F_T \equiv S_T$ it follows from the risk-neutral valuation principle that $\frac{E^Q[\log(F_T)]}{e^{r(T-t)}}$ is the price of an European style contingent claim (the log-contract) that pays the log value of the index at time T. Although this contract is not available on the market, its price can be inferred from the prices of the traded plain vanilla options. The classical result of Breeden and Litzenberger (1978) (see also Green and Jarrow(1987)) establishes the following link between the density $p(S)$ of the risk-neutral distribution of S_T at time T and the prices $C(S_t, K, T)$ of call options with maturity at time T and strike prices $K > 0$:

$$p\left(S\right)=e^{r(T-t)}\frac{\partial^{2}C\left(S_{t},K,T\right)}{\partial^{2}K}|_{K=S}. \quad (31)$$

The expressions (30) and (31) make possible the computation of the market expectations for the future variance $iv_{t,T}^2 = E^Q\left[\int_t^T \sigma_s^2 \partial s\right]$ under some relatively weak assumptions when a full spectrum of vanilla option prices is available.

B Appendix - ARFIMA Model Enhanced with the Implied Variance

One way to introduce the implied variance as an additional parameter in the AFRIMA model for the realized variance is to consider a specification of the following form. Let the log realized variance $\log(rv_t)$ follow an $ARFIMA\,(d,p,q)$ process. Set $\widetilde{y}_t = (1-L)^d \log(rv_t)$. Then $\widetilde{y}_t$ is an $ARMA\,(p,q)$ process that has the autoregressive specification:

$$\widetilde{y}_t = \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + u_t. \tag{32}$$

We can add additional regressors $\widetilde{\sigma}^2_{t-1}(bs), ..., \widetilde{\sigma}^2_{t-m}(bs)$ into (32) which account for the impact of the implied variance:

$$\widetilde{y}_t = \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + \gamma_1 \widetilde{\sigma}^2_{t-1}(bs) + ... + \gamma_m \widetilde{\sigma}^2_{t-m}(bs) + u_t \tag{33}$$

Additionally, let the MA part of the process $\widetilde{y}_t$ is given by:

$$u_t = \beta^*_1 u_{t-1} + ... + \beta^*_q u_{t-q} + \varepsilon_t. \tag{34}$$

By assumption, the residuals ε_t are uncorrelated with their lagged values and also uncorrelated with the lagged values of the dependent variable :

$$E\left[\varepsilon_t | \widetilde{y}_{t-1}, ..., \widetilde{y}_1, \widetilde{\sigma}^2_{t-1}(bs), ..., \widetilde{\sigma}^2_1(bs), \varepsilon_{t-1}, ..., \varepsilon_1\right] = 0. \tag{35}$$

To keep the model consistent with the use of the integrated series of the log realized variance, the parameters in the model that account for the implied variance are the lagged values of the integrated log implied variance $\log\left(\sigma^2_{t-i}(bs)\right), i = 1, 2, ...$. Indeed, it turns out that the log implied variance is fractionally integrated with the degree of fractional integration $d = 0.90$. The series $\widetilde{\sigma}^2_t(bs) = (1-L)^{d^*} \log\left(\sigma^2_t(bs)\right), d^* = 0.90$ is white noise. From (33) it follows that each residual u_{t-j} can be expressed as:

$$u_{t-j} = \widetilde{y}_{t-j} - \left(\alpha + \sum_{i=1}^{p} \beta_i \widetilde{y}_{t-i-j} + \sum_{i=1}^{m} \gamma_i \widetilde{\sigma}^2_{t-i-j}(bs)\right), \tag{36}$$

for $j = 1, 2, ...$ After substituting the last expression in (34), it follows :

$$u_t = \sum_{k=1}^{q} \beta_k^* \left(\widetilde{y}_{t-k} - \left(\alpha + \sum_{i=1}^{p} \beta_i \widetilde{y}_{t-i-k} + \sum_{i=1}^{m} \gamma_i \widetilde{\sigma}_{t-i-k}^2 (bs) \right) \right) + \varepsilon_t \tag{37}$$

$$= \sum_{k=1}^{q} \beta_k^* \widetilde{y}_{t-k} - \alpha \sum_{k=1}^{q} \beta_k^* - \sum_{k=1}^{q} \sum_{i=1}^{p} \beta_k^* \beta_i \widetilde{y}_{t-i-k} - \sum_{k=1}^{q} \sum_{i=1}^{m} \beta_k^* \gamma_i \widetilde{\sigma}_{t-i-k}^2 (bs) + \varepsilon_t.$$

Finally, the equation (33) can be written as:

$$\widetilde{y}_t = \alpha + \sum_{i=1}^{p} \beta_i \widetilde{y}_{t-i} + \sum_{i=1}^{m} \gamma_i \widetilde{\sigma}_{t-i}^2 (bs) + \sum_{k=1}^{q} \beta_k^* \widetilde{y}_{t-k} - \alpha \sum_{k=1}^{q} \beta_k^* \tag{38}$$

$$- \sum_{k=1}^{q} \sum_{i=1}^{p} \beta_k^* \beta_i \widetilde{y}_{t-i-k} - \sum_{k=1}^{q} \sum_{i=1}^{m} \beta_k^* \gamma_i \widetilde{\sigma}_{t-i-k}^2 (bs) + \varepsilon_t.$$

The last equation can be estimated as an non-linear regression using OLS. The condition (35) guarantees that the regression is correctly specified. Note that this approach is inconsistent with the methods for estimating the degree of fractional integration which usually assume that the process $\widetilde{y}_t = (1-L)^d \log (rv_t)$ is an $ARMA\,(p,q)$ process without the additional regressors $\widetilde{\sigma}_{t-1}^2 (bs), ..., \widetilde{\sigma}_{t-m}^2 (bs)$ in the autoregressive specification.
Another way of introducing implied variances which is used in the paper is the following. Consider the standard $ARFIMA\,(d,p,q)$model:

$$\widetilde{y}_t = (1-L)^d \log (rv_t) \tag{39}$$

$$\widetilde{y}_t = \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + u_t \tag{40}$$

$$u_t = \beta_1^* u_{t-1} + ... + \beta_q^* u_{t-q} + \varepsilon_t \tag{41}$$

$$E\left[\varepsilon_t | \widetilde{y}_{t-1}, ..., \widetilde{y}_1, \varepsilon_{t-1}, ..., \varepsilon_1\right] = 0. \tag{42}$$

The assumption (42) states that the residuals ε_t are not predictable when conditioned on the past information set represented by $\{ \widetilde{y}_{t-1}, \widetilde{y}_{t-1}, ..., \widetilde{y}_1, \varepsilon_{t-1}, ..., \varepsilon_1 \}$. However, they still can be predictable if they are conditioned on the past values of the implied variances. A simple way to introduce such predictability is to assume that the residuals ε_t can be represented as a linear combination of the lagged values of the fractionally integrated log implied variance plus a noise term

$$\varepsilon_t = \pi + \delta_1 \widetilde{\sigma}_{t-1}^2 (bs) + ... + \delta_m \widetilde{\sigma}_{t-m}^2 (bs) + \omega_t \tag{43}$$

where the residuals ω_t follow a MA process

$$\omega_t = \delta_1^* \omega_{t-1} + ... + \delta_l^* \omega_{t-l} + \eta_t \tag{44}$$

and new the residuals η_t are independent from the past values of the other variables in the model:

$$E\left[\eta_t | \widetilde{y}_{t-1}, ..., \widetilde{y}_1, \widetilde{\sigma}_{t-1}^2 (bs), ..., \widetilde{\sigma}_1^2 (bs), \eta_{t-1}, ..., \eta_1\right] = 0. \tag{45}$$

As before, this model can be rewritten as a non-linear regression. From expression (40) it follows

$$u_{t-j} = \widetilde{y}_{t-j} - \left(\alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t-k-j}\right) \tag{46}$$

for $i = 1, 2, ...$Combining (40) and (41) it follows:

$$\widetilde{y}_t = \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + \beta_1^* u_{t-1} + ... + \beta_q^* u_{t-q} + \varepsilon_t. \tag{47}$$

Using (46), each of the residuals $u_{t-1}, ..., u_{t-q}$ in (47) can be expressed as a sum of the past values of the dependent variable $\widetilde{y}_t$:

$$\widetilde{y}_t = \alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t-k} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t-i} - \left(\alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t-k-i}\right)\right) + \varepsilon_t = \tag{48}$$

$$= \alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t-k} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t-i} - \alpha\right) - \sum_{i=1}^{q} \sum_{k=1}^{p} \beta_i^* \beta_k \widetilde{y}_{t-k-i} + \varepsilon_t.$$

From the last expression follows

$$\varepsilon_{t-n} = \widetilde{y}_{t-n} - \left(\alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t-k-n} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t-i-n} - \alpha\right) - \sum_{i=1}^{q} \sum_{k=1}^{p} \beta_i^* \beta_k \widetilde{y}_{t-k-i-n}\right), \tag{49}$$

for $n = 1, 2, ...$ Similarly, the residuals $\omega_{t-1}, \omega_{t-2}, ...,$ can be written as a linear combination of the past values of the dependent variable $\widetilde{y}_t$. From (43) it follows:

$$\omega_{t-i} = \varepsilon_{t-i} - \left(\pi - \sum_{k=1}^{m} \delta_k \widetilde{\sigma}^2_{t-i-k} (bs) \right). \tag{50}$$

Substituting this expression into (44) yields:

$$\omega_{t-i} = \sum_{i=1}^{l} \delta_i^* \left(\varepsilon_{t-i} - \left(\pi - \sum_{k=1}^{m} \delta_k \widetilde{\sigma}^2_{t-i-k} (bs) \right) \right) + \eta_{t-i}. \tag{51}$$

The substitution of (49) into (51) will give a rather lengthy expression for the residuals $\omega_{t-1}, \omega_{t-2}, ...,$ as a linear combination of the past values of the dependent variable $\widetilde{y}_t$ plus the past values of the noise terms $\eta_{t-1,}\eta_{t-2,...,}$. Now the model (39)-(45) can readily written as a nonlinear regression. From (40)-(44) immediately follows:

$$\begin{gathered} \widetilde{y}_t = \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + u_t = \\ \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + \beta_1^* u_{t-1} + ... + \beta_q^* u_{t-q} + \varepsilon_t = \\ \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + \beta_1^* u_{t-1} + ... + \beta_q^* u_{t-q} \\ + \pi + \delta_1 \widetilde{\sigma}^2_{t-1} (bs) + ... + \delta_m \widetilde{\sigma}^2_{t-m} (bs) + \omega_t = \\ \alpha + \beta_1 \widetilde{y}_{t-1} + ... + \beta_p \widetilde{y}_{t-p} + \beta_1^* u_{t-1} + ... + \beta_q^* u_{t-q} + \pi + \delta_1 \widetilde{\sigma}^2_{t-1} (bs) + ... + \delta_m \widetilde{\sigma}^2_{t-m} (bs) \\ + \delta_1^* \omega_{t-1} + ... + \delta_l^* \omega_{t-l} + \eta_t. \end{gathered} \tag{52}$$

Substitution of the expressions (46) and (51) in place of the terms $u_{t-1}, ..., u_{t-q}$ and $\omega_{t-1}, ..., \omega_{t-l}$, respectively in (52) finally reduces the model (39)-(45) to a non-linear regression with the noise term η_t.

As it can be seen this model is analytically more elaborated than the first ARFIMA model extended with the implied variances. In addition, it is consistent with the methods for estimating the degree of fractional integration of the log realized variance, while the former model is not. Besides this purely theoretical objection for using the former model, estimation of this model shows that after including the values of $\widetilde{\sigma}^2_{t-1} (bs), ..., \widetilde{\sigma}^2_{t-m} (bs)$ into the autoregressive specification (33) for $\widetilde{y}_t$, the residuals ε_t in (34) are no longer uncorrelated.

The model also allows a convenient two-step estimation. First, the degree of fractional integration and the corresponding ARMA specification for the integrated series are estimated. Then the fitted residuals $\widetilde{\varepsilon}_t$ in (41) are regressed on the lagged values of the integrated implied variance assuming moving average error terms.

C Appendix - Calculation of the Multi-Step Predictions in the ARFIMA Model Enhanced with the Implied Variance

In the extended ARFIMA model the multi-step predictions are recursively computed in the following way. First the one-step prediction $\widetilde{y}_{t,t+1}$ of the integrated log realized variance $\widetilde{y}_t$ and the one-step prediction $\omega_{t,t+1}$ of the residual ω_t are calculated as follows. The formula (48) applied to $\widetilde{y}_{t+1}$ yields:

$$\widetilde{y}_{t+1} = \alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t+1-k} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t+1-i} - \alpha\right) - \sum_{i=1}^{q} \beta_i^* \sum_{k=1}^{p} \beta_k \widetilde{y}_{t+1-k-i} + \varepsilon_{t+1}. \tag{53}$$

Condition on the information set I_t available up to the day t the expectation of $\widetilde{y}_{t+1}$ is given by

$$\widetilde{y}_{t,t+1} = \alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t+1-k} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t+1-i} - \alpha\right) - \sum_{i=1}^{q} \beta_i^* \sum_{k=1}^{p} \beta_k \widetilde{y}_{t+1-k-i} + E\left[\varepsilon_{t+1} | I_t\right] \tag{54}$$

From (43) it follows that

$$E\left[\varepsilon_{t+1} | I_t\right] = \pi + \delta_1 \widetilde{\sigma}_t^2 \left(bs\right) + ... + \delta_m \widetilde{\sigma}_{t+1-m}^2 \left(bs\right) + E\left[\omega_{t+1} | I_t\right] \tag{55}$$

The expectation $\omega_{t,t+1} = E\left[\omega_{t+1} | I_t\right]$ equals:

$$\omega_{t,t+1} = \delta_1^* \omega_t + ... + \delta_l^* \omega_{t+1-l}. \tag{56}$$

Finally, the past residuals $\omega_t, ..., \omega_{t+1-l}$ can be readily computed from the data available up to the day t. This concludes the computation of the one-step predictions.

Now let $\widetilde{y}_{t,t+1}, ..., \widetilde{y}_{t,t+n}$ and $\omega_{t,t+1}, ..., \omega_{t,t+n}$ be the $n-$days ahead predictions of $\widetilde{y}_t$ and ω_t, respectively. From (48) it follows:

$$\begin{aligned} \widetilde{y}_{t+n+1} = \alpha &+ \sum_{k=1}^{p} \beta_k \widetilde{y}_{t+n+1-k} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t+n+1-i} - \alpha\right) \\ &- \sum_{i=1}^{q} \beta_i^* \sum_{k=1}^{p} \beta_k \widetilde{y}_{t+n+1-k-i} + \varepsilon_{t+n+1}. \end{aligned} \tag{57}$$

The $n+1$-day prediction $\widetilde{y}_{t,t+n+1} = E\left[\widetilde{y}_{t,t+n+1}|I_t\right]$ of $\widetilde{y}_t$ is given by

$$\begin{aligned}\widetilde{y}_{t,t+n+1} &= \alpha + \sum_{k=1}^{p} \beta_k \widetilde{y}_{t,t+n+1-k} + \sum_{i=1}^{q} \beta_i^* \left(\widetilde{y}_{t,t+n+1-i} - \alpha\right) \\ &- \sum_{i=1}^{q} \beta_i^* \sum_{k=1}^{p} \beta_k \widetilde{y}_{t,t+n+1-k-i} + E\left[\varepsilon_{t+n+1}|I_t\right], \end{aligned} \tag{58}$$

where $\widetilde{y}_{t,t-v} = \widetilde{y}_{t-v}$ if $v \geq 0$. The expectation $E\left[\varepsilon_{t+n+1}|I_t\right]$ equals:

$$\begin{aligned}E\left[\varepsilon_{t+n+1}|I_t\right] = \pi + \delta_1 E\left[\widetilde{\sigma}^2_{t+n+1-1}(bs)\,|I_t\right] + ... + \delta_m E\left[\widetilde{\sigma}^2_{t+n+1-1-m}(bs)\,|I_t\right] \\ + E\left[\omega_{t+n+1}|I_t\right]. \end{aligned} \tag{59}$$

As we noted before, the series $\widetilde{\sigma}^2_t(bs)$ is white noise, so the terms $E\left[\widetilde{\sigma}^2_{t+n+1-1}(bs)\,|I_t\right], ..., E\left[\widetilde{\sigma}^2_{t+n+1-1-m}(bs)\,|I_t\right]$ can be readily calculated as follows:

$$E\left[\widetilde{\sigma}^2_v(bs)\,|I_t\right] = \begin{cases} 0 & \text{,if } v \geq t+1 \\ \widetilde{\sigma}^2_v & \text{,if } v \leq t. \end{cases} \tag{60}$$

Finally, from (44) the prediction $\omega_{t,t+n+1} = E\left[\omega_{t+n+1}|I_t\right]$ for the residual ω_{t+n+1} can be represented as a linear combination of the past predicted values: $\omega_{t,t+n+1} = \delta_1^* \omega_{t,t+n} + ... + \delta_l^* \omega_{t,t+n-l}$. This concludes the calculation of the $n+1-$days ahead forecasts.

Table 1

Summary Statistics for the Realized Variance, Squared Returns and the Implied Variance

	Realized Variance	Squared Returns	Implied Variance
Mean	2.67E-04	2.84E-04	1.90E-04
Min	3.76E-05	0.000	0.001
Max	3.78E-03	0.015	7.37E-05
St.Dev.	3.25E-04	6.99E-04	1-23E-04
Autocorrelation at Different Lags[1]			
Lag 1	0.562	0.118	0.933
Lag 2	0.486	0.256	0.891
Lag 3	0.456	0.118	0.865
Lag 4	0.432	0.098	0.827
Lag 5	0.409	0.110	0.793
Lag 6	0.382	0.193	0.758
Lag 7	0.373	0.080	0.731
Lag 8	0.387	0.274	0.692
Lag 9	0.328	0.043	0.659
Lag 10	0.320	0.096	0.628

[1] The p-values for the corrsponding Ljung-Box Q-statistics are not reported since they show significance at a very high level (0.1%) for all three variables across all lags.

Table 2

In-Sample Estimation of the Different Volatility Models

	Regressor Term				
Model	C	ARCH	GARCH	IV	Log-likelihood
Daily GARCH	7.96E-06 (3.61E-06)**	0.092 (0.029)***	0.881 (0.032)***	-----	2487.691
Daily GARCH with IV	-4.89E-05 (2.63E-05)*	-0.065 (0.009)***	0.025 (0.297)	1.728 (0.542)***	2520.289
GARCH-RV	9.05E-06 (2.87E-06)***	0.287 (0.039)***	0.676 (0.041)***	-----	2528.337
GARCH-RV with IV	-3.29E-06 (4.11E-06)	0.241 (0.043)***	0.633 (0.051)***	0.198 (0.074)***	2528.407

Table 3

Summary Statistics for the Scaled Realized Variances

	GARCH-RV	GARCH-RV with IV
Mean	1.014	1.002
Min	0.103	0.113
Max	8.046	7.477
St.Dev	0.651	0.618
Autocorrelation at Different Lags[1]		
Lag 1	0.012	0.023
	(0.710)	(0.495)
Lag2	-0.040	-0.036
	(0.456)	(0.433)
Lag 3	-0.041	-0.034
	(0.376)	(0.441)
Lag 4	-0.026	-0.026
	(0.444)	(0.506)
Lag 5	0.036	0.033
	(0.424)	(0.503)
Lag 6	-0.018	-0.017
	(0.514)	(0.595)
Lag 7	-0.036	-0.027
	(0.493)	(0.626)
Lag 8	0.056	0.061
	(0.317)	(0.371)
Lag 9	0.000	-0.001
	(0.409)	(0.468)
Lag 10	0.033	0.035
	(0.413)	(0.459)

[1] The numbers in the brackets are the p-values for the corrsponding Ljung-Box Q-statistics.

Table 4

Regressing the In-Sample Outputs of the Different Models on each other

	Model											
	Daily GARCH			Daily GARCH with IV			GARCH-RV			GARCH-RV with IV		
Model	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2
Daily GARCH	-----	-----	-----	8.19E-05 (2.65E-05)***	0.664 (0.105)***	0.50	3.72E-05 (1.70E-05)**	0.807 (0.071)***	0.64	4.86E-05 (2.17E-05)**	0.783 (0.089)***	0.57
Daily GARCH with IV	7.71E-05 (1.34E-05)***	0.754 (0.058)***	0.50	-----	-----	-----	1.41E-05 (9.17E-06)	0.928 (0.039)***	0.75	-1.21E-05 (7.15E-06)*	1.045 (0.030)***	0.89
GARCH-RV	1.34E-05 (1.61E-05)***	0.791 (0.074)***	0.64	5.67E-05 (1.06E-05)***	0.803 (0.448)***	0.76	-----	-----	0.62	1.17E-05 (7.76E-06)**	0.973 (0.035)***	0.90
GARCH-RV with IV	8.45E-05 (1.46E-05)***	0.727 (0.068)***	0.56	3.91E-05 (6.90E-06)***	0.854 (0.031)	0.89	1.64E-05 (6.71E-06)**	0.920 (0.030)***	0.90	-----	-----	-----

Table 5

Regressions of the Realized Variance on its Predicted Values and the HRMSE errors
One-day Forecasting Horizon

	Model						
	Daily GARCH	Daily GARCH with IV	GARCH-RV	GARCH-RV with IV	ARFIMA	ARFIMA with IV	Linear Regression
Intercept	9.06E-05	5.21E-06	2.44E-05	2.14E-05	-2.97E-05	-4.87E-06	-8.97E-06
	(5.79E-05)	(2.76E-05)	(2.35E-05)	(2.03E-05)	(2.14E-05)	(3.58E-05)	(2.22E-05)
Slope	0.744	1.080	0.956	0.984	1.339	1.220	1.126
	(246)***	(0.121)***	(0.103)***	(0.091)***	(0.115)***	(0.159)***	(0.102)
R^2	0.25	0.48	0.44	0.47	0.48	0.48	0.51
HRMSE	0.79	0.59	0.64	0.60	0.55	0.53	0.56

Table 6

The Prediction Performance of the Different Volatility Models over Two to Ten Days Forecasting Horizon*

	HRMSE Criterion Forecasting Horizon								
	2	3	4	5	6	7	8	9	10
Daily GARCH	0.64	0.60	0.59	0.57	0.57	0.56	0.56	0.56	0.56
GARCH-RV	0.52	0.49	0.47	0.47	0.47	0.47	0.47	0.46	0.46
ARFIMA	0.39	0.37	0.36	0.36	**0.36**	0.36	0.37	0.37	0.38
ARFIMA with IV	**0.38**	**0.36**	**0.35**	**0.35**	**0.36**	0.36	0.36	0.37	0.38
Linear Regression	0.40	0.38	0.37	0.36	**0.36**	**0.35**	**0.35**	**0.36**	**0.36**

	R^2 Criterion Forecasting Horizon								
	2	3	4	5	6	7	8	9	10
Daily GARCH	0.25	0.25	0.25	0.25	0.25	0.23	0.22	0.21	0.20
GARCH-RV	0.44	0.47	0.51	0.52	0.52	0.51	0.49	0.46	0.43
ARFIMA	0.55	0.58	0.62	0.63	0.63	0.61	0.59	0.56	0.53
ARFIMA with IV	0.56	**0.61**	**0.64**	**0.64**	**0.64**	**0.64**	**0.62**	**0.60**	**0.57**
Linear Regression	**0.58**	**0.61**	0.63	0.63	**0.64**	**0.64**	**0.62**	**0.60**	**0.57**

*with bold are marked the models with the best performance for the corresponding forecasting horizon

Table 7

Regressions of the Realized Variance on its Predicted Values

	Forecasting Horizon														
	2 days			3 days			4 days			5 days			6 days		
Model	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2
Daily GARCH	2.13E-04 (2.44-04)	0.703 (0.266)***	0.25	3.36E-04 (1.82E-04)*	0.688 (0.2634)***	0.25	4.64E-04 (2.44E-04)*	0.677 (0.265)***	0.25	6.02E-04 (2.98E-04)**	0.664 (0.261)***	0.25	7.46E-04 (3.47E-04)**	0.650 (0.254)***	0.24
GARCH-RV	8.46E-05 (6.32E-05)	0.911 (0.142)***	0.44	1.17E-04 (1.06E-04)	0.931 (0.158)***	0.47	1.19E-04 (1.68E-04)	0.975 (0.187)***	0.51	1.22E-04 (2.55E-04)	1.002 (0.224)***	0.52	1.69E-04 (3.22E-04)	0.991 (0.235)***	0.52
ARFIMA	-7.76E-05 (5.60E-05)	1.410 (0.152)***	0.55	-1.52E-04 (9.66E-05)	1.489 (0.176)***	0.58	-2.79E-04 (1.39E-04)**	1.600 (0.194)***	0.62	-4.50E-04 (2.09E-04)**	1.716 (0.231)***	0.63	-5.97E-04 (2.86E-04)**	1.781 (0.262)***	0.63
ARFIMA with IV	-4.20E-05 (6.49E-05)	1.315 (0.153)***	0.56	-1.21E-04 (9.66E-05)	1.418 (0.146)***	0.61	-2.26E-04 (1.20E-04)*	1.509 (0.161)***	0.64	-3.45E-04 (1.70E-04)**	1.585 (0.183)***	0.64	-5.43E-04 (2.51E-04)**	1.700 (0.225)***	0.64
Linear Regression	-8.42E-06 (4.59E-05)	1.122 (0.110)***	0.58	-6.35E-06 (7.04E-05)	1.123 (0.115)	0.61	-2.26E-05 (9.47E-05)	1.151 (0.119)***	0.63	-4.84E-05 (1.28E-04)	1.181 (0.128)***	0.63	-1.19E-04 (1.87E-04)	1.233 (0.154)***	0.64

	Forecasting Horizon											
	7 days			8 days			9 days			10 days		
Model	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2
Daily GARCH	9.19E-04 (3.81-04)**	0.621 (0.239)***	0.23	1.10E-03 (4.14E-04)***	0.598 (0.226)***	0.22	1.268E-03 (4.53E-04)***	0.580 (0.218)***	0.21	1.46E-03 (4.92E-04)***	0.559 (0.211)***	0.20
GARCH-RV	2.47E-04 (3.75E-04)	0.965 (0.234)***	0.51	3.46E-04 (4.23E-04)	0.934 (0.230)***	0.49	4.75E-04 (4.66E-04)	0.899 (0.225)***	0.51	6.222E-04 (5.19E-04)	0.866 (0.225)***	0.43
ARFIMA	-6.95E-04 (3.37E-04)**	1.795 (0.269)***	0.61	-7.85E-04 (3.87E-05)**	1.804 (0.274)***	0.59	-8.51E-04 (4.33E-04)**	1.801 (0.276)***	0.56	-9.01E-04 (4.77E-04)*	1.791 (0.276**	0.53
ARFIMA with IV	-6.78E-04 (3.07E-04)**	1.740 (0.238)***	0.64	-7.91E-04 (3.58E-05)**	1.760 (0.246)***	0.62	-8.80E-04 (4.04E-04)**	1.765 (0.249***	0.60	-9.47E-04 (4.47E-04)**	1.759 (0.250)***	0.57
Linear Regression	-1.32E-04 (2.36E-04)	1.235 (0.167)***	0.63	-1.25E-04 (2.76E-04)	1.230 (0.172)	0.62	-1.30E-04 (3.10E-04)	1.233 (0.173)***	0.60	-1.30E-05 (3.47E-04)	1.238 (0.175)***	0.57

Table 8

Regressions for the Predicted Values on the Realized Variance

	Forecasting Horizon														
	2 days			3 days			4 days			5 days			6 days		
Model	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2
Daily GARCH	2.13E-04	0.703	0.25	3.36E-04	0.688	0.25	4.64E-04	0.677	0.25	6.02E-04	0.664	0.25	7.46E-04	0.650	0.24
	(2.44-04)	(0.266)***		(1.82E-04)*	(0.2634)***		(2.44E-04)*	(0.265)***		(2.98E-04)**	(0.261)***		(3.47E-04)**	(0.254)***	
GARCH-RV	8.46E-05	0.911	0.44	1.17E-04	0.931	0.47	1.19E-04	0.975	0.51	1.22E-04	1.002	0.52	1.69E-04	0.991	0.52
	(6.32E-05)	(0.142)***		(1.06E-04)	(0.158)***		(1.68E-04)	(0.187)***		(2.55E-04)	(0.224)***		(3.22E-04)	(0.235)***	
ARFIMA	-7.76E-05	1.410	0.55	-1.52E-04	1.489	0.58	-2.79E-04	1.600	0.62	-4.50E-04	1.716	0.63	-5.97E-04	1.781	0.63
	(5.60E-05)	(0.152)***		(9.66E-05)	(0.176)***		(1.39E-04)**	(0.194)***		(2.09E-04)**	(0.231)***		(2.86E-04)**	(0.262)***	
ARFIMA with IV	-4.20E-05	1.315	0.56	-1.21E-04	1.418	0.61	-2.26E-04	1.509	0.64	-3.45E-04	1.585	0.64	-5.43E-04	1.700	0.64
	(6.49E-05)	(0.153)***		(9.66E-05)	(0.146)***		(1.20E-04)*	(0.161)***		(1.70E-04)**	(0.183)***		(2.51E-04)**	(0.225)***	
Linear Regression	-8.42E-06	1.122	0.58	-6.35E-06	1.123	0.61	-2.26E-05	1.151	0.63	-4.84E-05	1.181	0.63	-1.19E-04	1.233	0.64
	(4.59E-05)	(0.110)***		(7.04E-05)	(0.115)		(9.47E-05)	(0.119)***		(1.28E-04)	(0.128)***		(1.87E-04)	(0.154)***	

	Forecasting Horizon											
	7 days			8 days			9 days			10 days		
Model	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2	intercept	slope	R^2
Daily GARCH	9.19E-04	0.621	0.23	1.10E-03	0.598	0.22	1.268E-03	0.580	0.21	1.46E-03	0.559	0.20
	(3.81-04)**	(0.239)***		(4.14E-04)***	(0.226)***		(4.53E-04)***	(0.218)***		(4.92E-04)***	(0.211)***	
GARCH-RV	2.47E-04	0.965	0.51	3.46E-04	0.934	0.49	4.75E-04	0.899	0.51	6.222E-04	0.866	0.43
	(3.75E-04)	(0.234)***		(4.23E-04)	(0.230)***		(4.66E-04)	(0.225)***		(5.19E-04)	(0.225)***	
ARFIMA	-6.95E-04	1.795	0.61	-7.85E-04	1.804	0.59	-8.51E-04	1.801	0.56	-9.01E-04	1.791	0.53
	(3.37E-04)**	(0.269)***		(3.87E-05)**	(0.274)***		(4.33E-04)**	(0.276)***		(4.77E-04)*	(0.276**	
ARFIMA with IV	-6.78E-04	1.740	0.64	-7.91E-04	1.760	0.62	-8.80E-04	1.765	0.60	-9.47E-04	1.759	0.57
	(3.07E-04)**	(0.238)***		(3.58E-05)**	(0.246)***		(4.04E-04)**	(0.249***		(4.47E-04)**	(0.250)***	
Linear Regression	-1.32E-04	1.235	0.63	-1.25E-04	1.230	0.62	-1.30E-04	1.233	0.60	-1.30E-05	1.238	0.57
	(2.36E-04)	(0.167)***		(2.76E-04)	(0.172)		(3.10E-04)	(0.173)***		(3.47E-04)	(0.175)***	