VOLATILITY FORECASTING FOR CRUDE OIL FUTURES

MASSIMILIANO MARZO*

UNIVERSITÀ DI BOLOGNA AND JOHNS HOPKINS UNIVERSITY

PAOLO ZAGAGLIA†

BI NORWEGIAN SCHOOL OF MANAGEMENT AND STOCKHOLM UNIVERSITY

This version: June 21, 2007

Abstract

This paper studies the forecasting properties of linear GARCH models for closing-day futures prices on crude oil, first position, traded in the New York Mercantile Exchange from January 1995 to November 2005. In order to account for fat tails in the empirical distribution of the series, we compare models based on the normal, Student's t and Generalized Exponential distribution. We focus on out-of-sample predictability by ranking the models according to a large array of statistical loss functions. The results from the tests for predictive ability show that the GARCH-G model fares best for short horizons from one to three days ahead. For horizons from one week ahead, no superior model can be identified. We also consider out-of-sample loss functions based on Value-at-Risk that mimic portfolio managers and regulators' preferences. EGARCH models display the best performance in this case.

JEL CLASSIFICATION: C22, G19.
KEYWORDS: GARCH models, kurtosis, oil prices, forecasting.

---

*Department of Economics, Università di Bologna, Piazza Scaravilli 2; 40126 Bologna, Italy; Johns Hopkins University, SAIS-BC. E-mail: marzo@economia.unibo.it. Web: http://www.dse.unibo.it/marzo/marzo.htm

†Department of Economics, BI Norwegian School of Management, Nydalsveien 37; 0484 Oslo, Norway; Department of Economics, Stockholm University, Universitetsvägen 10A; SE-106 91 Stockholm, Sweden. E-mail: pzaga@ne.su.se. Web: http://www.ne.su.se/~pzaga

"*The swings in oil prices that gave investors and traders whiplash in 2004 are not preventing new investors from rushing into oil and other energy-related commodities this year.* (...)

Ultimately, the rising number of speculators could lead to even more price volatility in 2005, pushing the highs higher and the lows lower. (...)

"*After a generation in the wilderness, the oil futures that are used to make a bet on oil prices have become a bona fide investment,*" *said Charles O'Donnell, who manages Lake Asset Management, a small energy fund based in London.*"

Heather Timmons, The New York Times[1]

1 INTRODUCTION

Futures contracts are one of the key instruments used to trade oil products in international financial markets (see Fleming and Ostdiek, 1999). Hence, the evolution of the daily volatility of oil futures prices conveys key information for understanding the functioning of oil markets.

Various studies analyze the usefulness of volatility models for the prediction of oil prices. In particular, Sadorsky (2006) considers univariate, bivariate and state-space models. He finds that single-equation GARCH overperforms more sophisticated models for forecasting petroleum futures prices. Fong and See (2002) study a Markov switching model of the cnditional volatility of crude oil futures prices, and show that the regimes identifies by their model capture major oil-related events. A related strand of literature investigates the transmission of volatility between energy markets. For instance, Ewing and Ozfidan (2002) show that there are significant patterns of volatility spillovers between the markets for oil and natural gas.

This paper evaluates the predictive performance of linear GARCH models for closing-day futures prices on crude oil traded in the New York Mercantile Exchange. In order to account for fat tails typical of financial series (see Bollerslev, 1987), we compare models based on the normal, Student's t and Generalized Exponential distribution. We focus on out-of-sample predictability over short (one to three days ahead) and long (one to three weeks ahead). Our empirical application ranks the models according to a large array of statistical loss functions.

The results from the tests for predictive ability show that the GARCH-G model fares best for short horizons from one to three days ahead. For horizons from one week ahead, no superior model can be identified. We also consider out-of-sample loss functions based on Value-at-Risk. Following Marcucci (2005), we introduce VaR-based functions that mimic portfolio managers and regulators' preferences for penalizing large forecast failures, as well as opportunity costs from over-investments. In this case, models of the EGARCH type display the best performance, followed closely by the GARCH-G.

The outline of the paper is as follows. Section 2 proposes an overview of univariate GARCH models. Section 3 outlines the forecast evaluation methods, including the statistical loss functions used in the paper, the tests for predictive ability and the Value-at-Risk strategies. Section 4 presents the dataset. The results are discussed in section 5. Section 6 proposes some concluding

---

[1] "*Real money pumps up volatility of oil prices*", January 7 2005.

remarks.

2   AN OVERVIEW OF GARCH MODELS

Let the model for the conditional mean of the return r_t take the form

$$r_t = \nu + \eta_t \sqrt{h_t} \tag{1}$$

where η_t is and i.i.d. process with variance h_t. In the standard GARCH(1,1) model, the model for the conditional variance is

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \nu h_{t-1} \tag{2}$$

with $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\nu_1 \geq 0$ in order to ensure a positive conditional variance. The presence of skewness in financial data has motivated the introduction of the Exponential GARCH (EGARCH) model:

$$\log(h_t) = \alpha_0 + \alpha_1 \left| \frac{\epsilon_{t-1}}{h_{t-1}} \right| + \gamma \frac{\epsilon_{t-1}}{h_{t-1}} + \nu \log(h_{t-1}) \tag{3}$$

The GJR model, instead, deals with the asymmetric reaction of the conditional variance depending on the sign of the shock:

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \left[1 - \mathcal{I}_{\{\epsilon_{t-1}>0\}} \right] + \gamma \epsilon_{t-1}^2 \mathcal{I}_{\{\epsilon_{t-1}>0\}} + \nu h_{t-1} \tag{4}$$

Bollerslev (1987) shows that financial time series are typically characterized by high kurtosis. In order to model the fat tails of the empirical distribution of the returns, we assume that the error term ϵ_t follows either a Student's t distribution with v degrees of freedom or a Generalized Error Distribution . The probability density function of ϵ_t then takes the form

$$f(\epsilon_t) = \frac{\Gamma((v+1)/2)}{\sqrt{\pi}\Gamma(v/2)} (v-2)^{-1/2} (h_t)^{-1/2} \left[1 + \frac{\epsilon_t^2}{h_t(v-2)} \right]^{-\frac{v+1}{2}} \tag{5}$$

where $\Gamma(\cdot)$ indicates the Gamma function with the shape parameter $v > 2$. Under the Generalized Error Distribution (G), the model errors follow the pdf

$$f(\epsilon_t) = \frac{v \exp\left(1/2 \left| \frac{\epsilon_t}{\lambda h_t^{1/2}} \right|^v \right)}{h_t^{1/2} \lambda 2^{(2+1/v)} \Gamma(1/v)} \tag{6}$$

with $\lambda := \left[\left(2^{-2/v} \Gamma(1/v) \right) / \Gamma(3/v) \right]^{1/2}$.

3   FORECAST EVALUATION

The $m-$step ahead volatility forecast, indicated by $\hat{m}_{t,t+m}$, is computed as the aggregated sum of the forecasts for the following m steps made at time t. We consider the volatility forecast over three horizons, namely one day, one week and three weeks ahead.

3.1   STATISTICAL LOSS FUNCTIONS

There exists no unique criterion capable of selecting the 'best' forecasting model. Hence, this paper evaluates the predictive performance of the GARCH models through an array of statistical loss functions. These criteria are listed in table 4. The functions named MSE_1 and MSE_2 are typical mean squared error metrics. The $R2LOG$ function penalizes the volatility forecasts for low volatility periods in a way different from high volatility periods. Finally, the Mean Absolute Deviation (MAD) criterion is robust to the presence of outliers. Bollerslev and Ghysels (1996) have proposed the heteroskedasticity-adjusted MSE ($HMSE$).

It is instructive to report the so-called Success Ratio (SR). This statistics indicates the fraction of the volatility forecasts that have the same direction of change as the realized volatility. For an actual volatility proxy $\bar{\sigma}_{t+m}$ at time $t+m$, and a volatility forecast $\bar{h}_{t,t+m}$, the SR can be written as

$$SR := (1/n) \sum_{j=0}^{n-1} \mathcal{I}_{\{\bar{\sigma}_{t+m+j}\bar{h}_{t+j,t+m+j}>0\}} \tag{7}$$

where $\mathcal{I}$ is an indicator function.

The Directional Accuracy (DA) test of Pesaran and Timmermann (1992) is based on the statistics

$$DA := \frac{SR - SRI}{\sqrt{var(SR) - var(SRI)}} \tag{8}$$

where $SRI := P\hat{P} + (1-P)(1-\hat{P})$, P indicates the fraction of times such that $\bar{\sigma}_{t+m+j}$, and $\hat{P}$ is the fraction of times for which $\bar{h}_{t+m+j} > 0$.

3.2   TESTS OF PREDICTIVE ABILITY

Diebold and Mariano (1995) propose a test of equal predictive ability between two competing models. Denote by $\{e_{i,t}\}_{t=1}^n$ and $\{e_{j,t}\}_{t=1}^n$ the forecast errors of two models i and j. The loss differential between the two forecasts can be written as

$$d_t := [g(e_{i,t}) - g(e_{j,t})] \tag{9}$$

where $g(\cdot)$ is the loss function. If $\{d_t\}_{t=1}^n$ is covariance stationary and has no long memory, the sample mean loss differential $\bar{d} = (1/n)\sum_{t=1}^n d_t$ is asymptotically distributed as $\sqrt{n}(\bar{d} - \mu) \xrightarrow{d} N\left(0, V(\bar{d})\right)$. Under the null of equal predictive ability, Diebold and Mariano (1995) propose the test statistics $DM := \bar{d}/\sqrt{\hat{V}(\bar{d})} \sim N(0,1)$. Harvey, Leybourne, and Newbold (1997) suggest a

Modified DM statistics (MDM) that tackles the oversize problem that arises in small samples. The modified test statistics is obtained by multiplying the standard statistics by a factor of correction.

White (1980) introduces a test for superior predictive ability — the RC test — that checks whether a specific forecasting model is outperformed by an alternative set of models according to a loss function. Let $\mathcal{L}\left(\hat{\sigma}_t^2, \hat{h}_{k,t}\right)$ denote the loss function for the prediction with model k, with $k = 1, \ldots l$. The relative predictive performance of model 0 can be computed as

$$f_{k,t} = \mathcal{L}_{t,0} - \mathcal{L}_{t,k} \tag{10}$$

If $f_{k,t}$ is stationary, we can define the expected relative performance $E\left[f_{k,t}\right]$. The testing procedure amounts to checking that none of the competing models outperforms the benchmark:

$$\mathcal{H}_0 : \max_{k=1,\ldots l} E\left[f_{k,t}\right] \leq 0 \tag{11}$$

The rejection of the null implies that at least one competing model is better than the benchmark. The test statistics is

$$\max_{k=1,\ldots l} n^{1/2} \bar{f}_{k,n} \tag{12}$$

Hansen (2005) stresses that the distribution of the test statistics is not unique under the null, and that it is sensitive to the inclusion of poor models. Hence, he proposes a way of obtaining a consistent estimate of the $p-$value of a modified test statistics, along with an upper and a lower bound. The resulting SPA_u yields the $p-$value of a conservative test where all the competing models are assumed to be as good as the benchmark in terms of expected loss. The SPA_l test is instead based on $p-$values that assume that the models with bad performance are poor models.

3.3   VALUE-AT-RISK

As suggested by Brooks and Persand (2003), loss functions based on Value-at-Risk are a natural alternative to the standard statistical loss functions while evaluating the predictive performance of a model estimated on financial data. The VaR measures the market risk of a portfolio quantified in monetary terms, and arising from market fluctuations at a given significance level.

Several statistical tests can be computed to assess the forecasting ability of the GARCH models for the VaR. The Time Until First Failure ($TUFF$) is based on the failure process, namely the number of exceptions of the VaR from model k — $\mathcal{I}_{r_t < \text{VaR}_t^k}$. For a significance level γ, the null hypothesis is $\mathcal{H}_0 : \gamma = \gamma_0$, and the likelihood-ratio test statistics is:

$$LR_{TUFF}(\tilde{T}, \hat{\gamma}) = -2 \log \left[\hat{\gamma}(1 - \hat{\gamma})^{\tilde{T}-1}\right] + 2 \log \left[\tilde{T}^{-1} \left(1 - \tilde{T}^{-1}\right)^{\tilde{T}-1}\right] \tag{13}$$

with the number of observations $\tilde{T}$ before the first exception. The statistics LR_{TUFF} is distributed as a $\chi^2(1)$ under the null. The 95% confidence interval is (3, 514) for the 99% VaR, and (1, 101) for the 95% VaR.

A VaR can be insufficient to cover the losses that a portfolio incurs. In this sense, a model can

be judged adequate when the proportion of failures out-of-sample is close to the nominal value. The unconditional criterion suggests that the VaR is adequate if $E[\mathcal{I}_t] = \gamma$. Since the number of failures is i.i.d. and distributed as a binomial, the likelihood-ratio test statistics can be written as

$$LR_{PF} = -2\log\left[\frac{\gamma^{n_1}(1-\gamma)^{n_0}}{\hat{\gamma}^{n_1}(1-\hat{\gamma})^{n_0}}\right] \sim \chi^2(1) \tag{14}$$

where n_1 is the number of failures, γ is the level of the VaR, and $\hat{\gamma} := n_1/(n_1 + n_0)$.

Since financial data are characterized by volatility clustering, good interval forecasts from a VaR model should be narrow in periods of low volatility, and wide in periods of high volatility. Christoffersen (1998) proposes a test of independence. The null of independent failure rates is tested against a first-order Markov failure process. The test statistics takes the form

$$LR_{IND} = -2\log\left[\frac{(1-\hat{\pi})^{(n_{00}+n_{10})}(1-\hat{\pi})^{(n_{01}+n_{11})}}{(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}}\right] \sim \chi^2(1) \tag{15}$$

where $\pi_{ij} = \Pr\{\mathcal{I}_t = i | \mathcal{I}_{t-1} = j\}$. Finally, we consider a conditional test of correct coverage where the null of independent failures with a probability γ is tested against the first-order Markov failure:

$$LR_{CC} = -2\log\left[\frac{(1-\gamma)_0^n\gamma^{n_{10}}}{(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}}\right] \sim \chi^2(2) \tag{16}$$

We follow Marcucci (2005) and evaluate the competing models through VaR loss functions that mimic the utility functions of risk managers. In particular, the Regulator Loss Fucntion (RLF) introduces an asymmetric penalty for the large losses. The RLF takes the form

$$\mathcal{L}_t^1 := (r_t - \text{VaR}_t^k)^2 \mathcal{I}_{\{r_t < \text{VaR}_t^k\}} \tag{17}$$

The Firm Loss Function (FLF), instead, penalizes the models that require an excessive investment of capital, and that bear larger opportunity costs. This function is defined as

$$\mathcal{L}_t^2 := (r_t - \text{VaR}_t^k)^2 \mathcal{I}_{\{r_t < \text{VaR}_t^k\}} - \nu \text{VaR}_t^k \mathcal{I}_{\{r_t > \text{VaR}_t^k\}} \tag{18}$$

4 DATA

The dataset consists of daily observations of closing-day futures prices on crude oil traded in the New York Mercantile Exchange. We focus on futures on the first position. The series spans from January 2 1995 to November 22 2005, for a total of 2842 observations. We use 2080 observations for in-sample analysis, and the remaining 762 for out-of-sample forecasts. The GARCH models are estimated on the percentage returns $r_t := 100\log(p_t/p_{t-1})$.

Figure 6 plots the data series. The plot of the empirical distribution with respect to the theoretical standard normal gives indication of fat tails due to extreme observations. Table 2 report the main properties of the data. The kurtosis coefficient is larger than 3, and supports the hypothesis of fat-tailed distribution. The Jarque-Bera statistics suggests that the returns are consistent with a strong deviation from normality. Table 2 includes the results from the the

normality test of Anderson and Darling (1952). This is a modification of the Kolmogorov-Smirnov test, and gives more weight to the tails than the Kolmogorov-Smirnov test itself. Also in this case, there is a rejection of the null of normality. The significance of the Ljung-Box statistics up to the twelfth order points towards the presence of ARCH effects in the returns (see table 2). Table 3 reports the results from the LM tests for the null of no ARCH of Engle (1982). Again, the null is rejected strongly.

The GARCH models are estimated through quasi-maximum likelihood by maximizing the log-likelihood function obtained as the logarithm of the product of the conditional densities of the prediction errors. The maximization step is carried out by the Broyden, Fletcher, Goldfarb and Shanno Newton algoritm.

A measure of 'true volatility' is required for the evaluation of the forecasting performace of the models. Andersen and Bollerslev (1998) suggest that using intra-daily returns removes most of the noise that arises from the use of daily data. However, intra-daily series are hardly obtainable for the type of futures considered in this paper. Hence, we approximate the true volatility through the actual volatility at each point in time.

5   RESULTS

5.1   ESTIMATED MODELS

The estimates of the parameters of the GARCH models are reported in table 1.[2] The standard errors are robustified again heteroskedasticity through a Sandwich formula. The first point of interest concerns the fact that not all the conditional means are statistically significant at standard confidence levels (e.g. see the EGARCH-G). Most of the parameters of the conditional variance retain statistical validity. For the models based on the t distribution, the conditional kurtosis is equal to $3(v-2)/(v-4)$. The resulting estaimtes of conditional kurtosis are all larger than 6 for all the specifications. This confirms the importance of modelling fat-tailed distributions for oil futures. Also the models based on the GED support the evidence for fat tails. In this case, the conditional kurtosis takes a value of $(\Gamma(1/v)\Gamma(5/v))/((\Gamma(1/v))^2)$, which gives 6.6127 for the GARCH-G, 6.3802 for the EGARCH-G, and 6.5451 for the GJR-G.

5.2   IN-SAMPLE FORECAST EVALUATION

Table 5 reports some descriptive statistics for in-sample evaluation. These tests can be used for model selection. The maximized log-likelihood suggests that the GJR model with t errors provides the most accurate description of the data. Also according to the Akaike and Schwartz information criteria, the GRJ-t model fits the best. However, there is no unique best alternative emerging from the use of the statistical loss functions of table 4. Except for the HMSE, the main pattern concerns the fact that the GARCH models estimated with $t-$distribution obtain the highest ranking. This suggests that the estimates are capable of capturing the leptokurtosis of the empirical distribution of the returns.

---

[2]Since the focus of this paper is on predictability and risk management, we do not conduct any specification test.

5.3 OUT-OF-SAMPLE FORECAST EVALUATION

A good in-sample fit provides no indication for the forecasting performance of a model out-of-sample. Tables 6 and 7 report the evaluation for out-of-sample forecasts over the short term (one, two and three days ahead), and over longer-term horizons (one, two and three weeks ahead). The proxy for the true volatility is the realized (daily) volatility. All but one the DA tests statistics are statistically significant. The GARCH-G model provides the best forecasts for one, two and three days ahead. For forecasts one-week ahead, the GARCH-G and EGARCH-G models are competitors. Instead, the EGARCH-G model is the best performer for two and three weeks ahead.

Tables 8-14 report both the results from the DM and modified DM tests. As benchmarks, we use the models that perform the best in the DM tests. Again, all the statistical loss functons of table 4 are used for the comparison. For short horizons, tables 8-10 use the GARCH-G as benchmark. The results indicate that the null of equal predictive ability is rejected strongly, suggesting that the benchmark outperforms the competing models. Furthermore, the sign of the tests statistics is negative, indicating that the loss is lower under the benchmark than under the alternative model for all pairwise comparisons.

Tables 11 and 12 consider one-week ahead forecasts, with the GARCH-G and EGARCH-G, respectively, compared to the other models. When the GARCH-G model is the benchmark, the EGARCH-G fares better for almost all the loss functions (see table 11). The reserve happens when the EGARCH-G is the benchmark (see table 12). Finally, for a predictive horizon of two and three weeks ahead, the EGARCH-G model does not outperform two models which do not rank well in terms of DM tests. Results not reported here suggest that these alternative competitors generate higher statistical losses when used as benchmark with respect to the EGARCH-G model. Overall, the GARCH-G appears to be the most appropriate model for short-term forecasts. At longer horizons, a suitable benchmark cannot be found.

Tables 15-17 report the results from the reality check and super-predictive ability for short horizons. Each model is evaluated against all the others. For every model, the rows indicate the p-values of the RC tests. SPA_l^0 and SPA_c^0 refer to the p-values of Hansen (2005) computed through a stationary bootstrap with 3000 re-samples. The main result concerns the fact that, when the GARCH-G is the benchmark, the null of SPA is not rejected for all the loss functions at short horizons. There are also occasional rejections when the GARCH-t and GJR-G are the benchmark, albeit with lower p-values. This results should not be striking as they obtained also by Marcucci (2005) on stock market data. For instance, Hansen and Lunde (2006) suggest that the GARCH(1,1) is not the best specification when compared with other models. The outcomes of the reality check over long horizons are reported in tables 18-20. Strinkingly, the GARCH-G model never rejects the null of SPA for all the loss functions independently from the predictive horizon. This is a relevant results, as it casts doubts on the lack of predictive power of the GARCH-G for long horizons emphasized by the DM tests. Occasional acceptances are also displayed by the GARCH-t, EGARCH-G and the GJR-G models.

In the following step, we compare the models with measures of conditional and unconditional coverage of VaR estimates. Following Marcucci (2005), we also introduce subjective loss functions that are meant to mimic the preferences of risk managers. The RLF and LFL penalize large failures in the VaR forecast. Tables 21 and 22 present the VaR estimates at the 95% and 99% for short and

long horizons, respectively. The table shows the results from the test of correct coverage (LR_{PF}) to check if PF is significantly higher than the nominal rate, the LR_{IND} test of independence, and the test of correct conditional coverage LR_{CC}. Numbers in bold identify the minima for each evaluation criterion. The theoretical TUFF at 5 and 1% are, respectively, 20 and 100.

At both short and long horizons, all the models but one display failures with respect to the theoretical TUFF for the 95% and 99% VaR. In terms of probability of failure, for short horizons, all the models with t distributed disturbances are inadequate for the 95% VaR, as they are rejected for a too high PF. However, there are no rejections for the GARCH-G model, which fares best in terms of statistical criteria of forecast evaluation. Table 21 shows that the three tests of correct unconditional and conditional coverage do not reject the GARCH-G. However, when the aim is that of covering 99% of losses, there are more models that perform equally well for each test of conditional and unconditional coverage. The last two columns of tables 21 and 22 report the average RLF and FLF. The GARCH-G model never yields the lowest values for short-horizon forecasts. However, for the 99% VaR, average losses closer to the lowest values are delivered. At long horizons, the GARCH-G model delivers better average RLF and FLF. An overall look at the results shows that EGARCH models — the EGARCH-t for short horizons and the EGARCH-G for long horizons — fare better than both GARCH and GJR models in terms of Value-at-Risk loss functions.

6 CONCLUSION

This paper studies the forecasting properties of linear GARCH models for closing-day futures prices on crude oil, first position, traded in the NYMEX. We compare volatility models based on the normal, Student's t and Generalized Exponential distribution. Our focus is on out-of-sample predictability. To that end, we rank the models according to a large array of statistical loss functions.

The results from the tests for predictive ability show that the GARCH-G model fares best for short horizons from one to three days ahead. For horizons from one week ahead, no superior model can be identified. We also consider out-of-sample loss functions based on Value-at-Risk that mimic portfolio managers and regulators' preferences for penalizing large forecast failures and opportunity costs from over-investments. In this case, EGARCH models exhibit the best performance, followed closely by the GARCH-G.

References

ANDERSEN, T. G., AND T. BOLLERSLEV (1998): "Answering the Critics: Yes, ARCH models Do Provide Good Volatility Forecasts," *International Economic Review*, 39.

ANDERSON, T. W., AND D. A. DARLING (1952): "Asymptotic Theory of Certain Goodness-of-Fit Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23.

BOLLERSLEV, T. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return," *The Review of Economics and Statistics*, 69.

BOLLERSLEV, T., AND E. GHYSELS (1996): "Periodic Autoregressive Conditional Heteroskedasticity," *Journal of Business and Economic Statistics*, 14.

BROOKS, C., AND G. PERSAND (2003): "Volatility Forecasting for Risk Management," *Journal of Forecasting*, 22.

CHRISTOFFERSEN, P. F. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50.

EWING, BRADLEY T., F. M., AND O. OZFIDAN (2002): "Volatility transmission in the oil and natural gas markets," *Energy Economics*, 24.

FLEMING, J., AND B. OSTDIEK (1999): "The impact of energy derivatives on the crude oil market," *Energy Economics*, 21.

FONG, W. M., AND K. H. SEE (2002): "A Markov switching model of the conditional volatility of crude oil futures prices," *Energy Economics*, 24.

HANSEN, P. R. (2005): "A Test for Superior Predictive Ability," *Journal of business and Economic Statistics*, 23.

HANSEN, P. R., AND A. LUNDE (2006): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?," *forthcoming in the Journal of Applied Econometrics*.

HARVEY, D., S. LEYBOURNE, AND P. NEWBOLD (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13.

MARCUCCI, J. (2005): "Forecasting Stock Market Volatility with Regime-Switching GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 9.

PESARAN, M. H., AND A. TIMMERMANN (1992): "A Simple Nonparametric Test of Predictive Performance," *Journal of Business and Economic Statistics*, 10.

SADORSKY, P. (2006): "Modeling and forecasting petroleum futures volatility," *Energy Economics*, 28.

WHITE, A. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48.
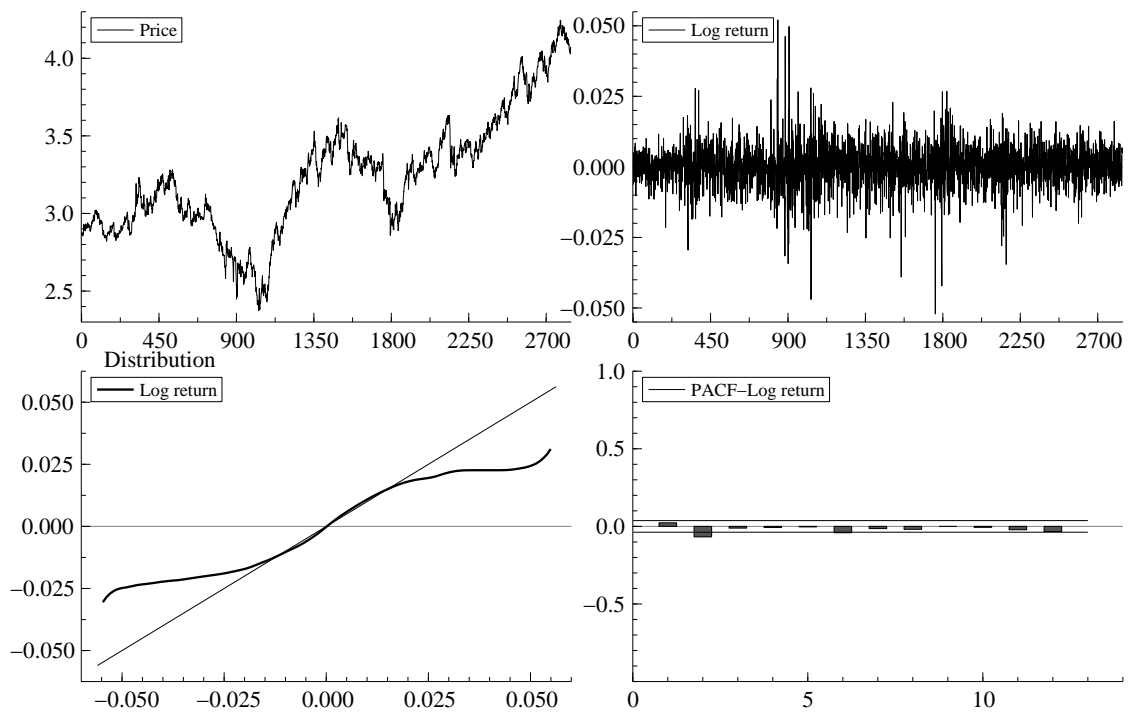
Figure 1: Futures prices for crude oil

Table 1: Estimates of GARCH models

	GARCH-N	GARCH-t	GARCH-G	EGARCH-N	EGARCH-t	EGARCH-G	GJR-N	GJR-t	GJR-G
ν	0.096 [0.051]	0.114 [0.047]	0.059 [0.044]	0.059 [0.052]	0.098 [0.048]	0.046 [0.044]	0.072 [0.052]	0.103 [0.048]	0.050 [0.044]
α_0	0.735 [0.156]	0.765 [0.308]	0.770 [0.303]	0.015 [0.015]	0.030 [0.028]	0.029 [0.030]	0.707 [0.136]	0.768 [0.272]	0.763 [0.267]
α_1	0.081 [0.011]	0.048 [0.016]	0.059 [0.017]	0.081 [0.013]	0.060 [0.022]	0.073 [0.025]	0.109 [0.015]	0.077 [0.023]	0.089 [0.025]
ν	0.794 [0.033]	0.822 [0.062]	0.808 [0.062]	−0.068 [0.008]	−0.053 [0.016]	−0.059 [0.016]	0.801 [0.028]	0.822 [0.055]	0.811 [0.055]
γ	-	-	-	0.956 [0.008]	0.956 [0.018]	0.952 [0.019]	0.046 [0.011]	0.013 [0.017]	0.024 [0.018]
v	-	5.245 [0.648]	1.199 [0.046]	-	5.345 [0.656]	1.206 [0.046]	-	5.299 [0.650]	1.201 [0.047]

Legend: Brackets report standard errors.

Table 2: Descriptive statistics of the returns

Max	32.43
Min	-37.57
Mean	6.7e-2
Standard dev.	3.82
Kurtosis	13.50
Skewness	-0.433
JB	1.7e+3*
Anderson-Darling	17.4557 [0.0]
LJB(12)	16.13*

Legend: Brackets report the marginal probability. The LJB(12) is the Ljung-Box test statistics on the squared residuals from the regression of the conditional mean. Under the null of no serial correlation, it is distributed as a $\chi^2(q)$ distribution with q lags. Like for the LM test, the critical value is 21.03. JB is the Jarque-Bera test of normality. It has a χ^2 distribution with 2 degrees of freedom. The critical value at the 5% level is 5.99.

Table 3: Tests of Engle (1982)

Lag	Engle (1982)
1	16.7311 [0.0]
2	29.9616 [0.0]
3	54.1671 [0.0]
4	59.3947 [0.0]
5	64.7837 [0.0]
6	64.9729 [0.0]
7	65.0260 [0.0]
8	65.4225 [0.0]
9	65.6834 [0.0]
10	66.5826 [0.0]

Legend: p-values are in brackets. This table resport the ARCH LM test statistics up to the twelfth lag. Under the null of no ARCH, it has a $\chi^2(q)$ distribution, with q the number of lags.

Table 4: Statistical loss functions

MSE_1	$1/n \sum_{t=1}^{n} \left(\hat{\sigma}_{t+m} - \hat{h}_{t,t+m}^{1/2} \right)^2$
MSE_2	$1/n \sum_{t=1}^{n} \left(\hat{\sigma}_{t+m}^2 - \hat{h}_{t,t+m} \right)^2$
$QLIKE$	$1/n \sum_{t=1}^{n} \left(\log \hat{h}_{t,t+m} + \hat{\sigma}^2 \hat{h}_{t,t+m}^{-1} \right)$
$R2LOG$	$1/n \sum_{t=1}^{n} \left(\log \left[\hat{\sigma}^2 \hat{h}_{t,t+m}^{-1} \right] \right)^2$
MAD_1	$1/n \sum_{t=1}^{n} \left| \hat{\sigma}_{t+m} - \hat{h}_{t,t+m}^{1/2} \right|$
MAD_2	$1/n \sum_{t=1}^{n} \left| \hat{\sigma}_{t+m}^2 - \hat{h}_{t,t+m} \right|$
$HMSE$	$1/n \sum_{t=1}^{n} \left(\hat{\sigma}^2 \hat{h}_{t,t+m}^{-1} - 1 \right)^2$

Table 5: In-sample predictability

Model	Pers	AIC	Rank	BIC	Rank	Log(L)	Rank	MSE_1	Rank	MSE_2	Rank	QLIKE	Rank	R2LOG	Rank	MAD_2	Rank	MAD_1	Rank	HMSE	Rank
GARCH-N	0.944	4.43	7	4.44	7	-4605.24	8	3.03	7	181.88	8	2.59	3	10.08	4	6.11	7	1.39	5	4.29	3
GARCH-t	0.974	4.43	6	4.44	6	-4601.20	6	2.97	2	179.52	1	2.59	1	10.06	2	6.00	3	1.38	2	4.65	7
GARCH-G	0.944	4.43	8	4.45	8	-4605.24	7	3.03	6	181.87	7	2.59	2	10.08	3	6.11	6	1.39	4	4.29	4
EGARCH-N	0.951	4.36	4	4.37	4	-4526.95	5	3.05	9	181.40	3	2.59	7	10.24	9	6.14	8	1.41	8	4.19	1
EGARCH-t	0.727	5.00	9	5.01	9	-5190.18	9	2.83	1	198.86	9	4.11	9	7.13	1	4.91	1	1.12	1	86.67	9
EGARCH-G	0.952	4.36	5	4.37	5	-4526.41	4	3.05	8	181.52	4	2.59	8	10.23	8	6.15	9	1.41	9	4.24	2
GJR-N	0.949	4.34	2	4.36	2	-4510.73	3	3.03	5	181.53	5	2.59	5	10.12	7	6.10	5	1.39	7	4.34	5
GJR-t	0.989	4.33	1	4.35	1	-4502.21	1	2.98	3	180.18	2	2.59	4	10.10	5	6.00	2	1.38	3	5.21	8
GJR-G	0.949	4.34	3	4.36	3	-4510.61	2	3.02	4	181.54	6	2.59	6	10.12	6	6.10	4	1.39	6	4.36	6

Table 6: Out-of-sample predictability: short horizons

Model	MSE1	Rank	MSE2	Rank	QLIKE	Rank	R2LOG	Rank	MAD2	Rank	MAD1	Rank	HMSE	Rank	SR	DA
					1-step ahead volatility forecast											
GARCH-N	2.5341	7	95.6489	9	2.6486	9	7.9849	4	1.3143	6	5.484	6	4.3277	9	0.58	0.0702
GARCH-t	2.2847	2	82.6025	3	2.4891	2	7.8479	2	1.274	2	5.2946	2	1.462	3	0.68	6.6520**
GARCH-G	2.2034	1	77.997	1	2.4614	1	7.8447	1	1.2535	1	5.1799	1	1.1987	1	0.71	8.1032**
EGARCH-N	2.584	9	95.0834	7	2.6299	7	8.1556	9	1.3402	9	5.595	9	3.3602	7	0.59	-0.2702
EGARCH-t	2.4975	6	89.2458	6	2.5531	6	8.1168	8	1.3316	8	5.5739	8	1.8151	6	0.65	4.2484**
EGARCH-G	2.3745	4	84.1185	4	2.5158	4	8.0643	7	1.3032	4	5.4091	4	1.4671	4	0.7	7.3769**
GJR-N	2.5528	8	95.2812	8	2.6361	8	8.0632	6	1.3259	7	5.5325	7	3.7637	8	0.57	-0.8742
GJR-t	2.4107	5	87.1772	5	2.5303	5	7.9888	5	1.3063	5	5.4563	5	1.7387	5	0.66	5.1195**
GJR-G	2.2966	3	81.8286	2	2.4935	3	7.9496	3	1.2799	3	5.3034	3	1.3813	2	0.71	7.7374**
					2-step ahead volatility forecast											
GARCH-N	2.7155	7	209.958	7	3.4977	9	1.9626	7	1.2127	7	7.4771	7	6.1283	9	0.69	7.9131**
GARCH-t	2.5142	4	202.3727	4	3.388	4	1.7873	1	1.1616	2	7.2489	2	3.9159	4	0.71	9.6224**
GARCH-G	2.3195	1	187.8972	1	3.2828	1	1.8069	2	1.1452	1	7.1388	1	2.5298	1	0.77	12.5998**
EGARCH-N	2.7319	8	210.987	8	3.4736	7	2.0072	9	1.2198	9	7.5161	9	5.4045	7	0.68	6.4727**
EGARCH-t	2.6514	6	208.2175	6	3.4287	6	1.9311	6	1.1939	6	7.3997	6	4.4123	6	0.66	6.2366**
EGARCH-G	2.4959	3	198.3374	3	3.3433	3	1.9134	5	1.1803	4	7.3189	4	3.1437	3	0.75	11.3323**
GJR-N	2.7355	9	211.1724	9	3.4905	8	1.988	8	1.2187	8	7.5097	8	5.7888	8	0.68	6.8351**
GJR-t	2.6114	5	206.6676	5	3.4224	5	1.8793	4	1.1842	5	7.355	5	4.331	5	0.69	7.9797**
GJR-G	2.4384	2	194.9228	2	3.3264	2	1.8703	3	1.1693	3	7.2628	3	2.9549	2	0.75	11.5059**
					3-step ahead volatility forecast											
GARCH-N	4.1234	5	381.6426	4	4.3586	6	1.3623	5	1.5106	6	11.0936	6	9.1091	6	0.75	11.8736**
GARCH-t	4.3898	9	392.5032	9	4.5147	9	1.4666	9	1.5757	9	11.3899	9	9.7863	9	0.72	10.8861**
GARCH-G	3.6918	1	357.0735	1	4.1252	1	1.1984	1	1.4428	1	10.7525	1	5.2089	1	0.8	14.9482**
EGARCH-N	4.0986	4	382.9031	5	4.3238	4	1.3458	4	1.4947	4	11.0301	4	8.5937	4	0.72	9.8668**
EGARCH-t	4.231	7	389.1489	7	4.409	7	1.3912	7	1.5183	7	11.1318	7	9.274	7	0.67	6.9296**
EGARCH-G	3.85	3	369.4634	3	4.1845	3	1.25	3	1.4565	2	10.8392	2	6.1513	3	0.78	13.5538**
GJR-N	4.1373	6	383.5959	6	4.3545	5	1.3641	6	1.5074	5	11.0859	5	8.9972	5	0.73	10.1970**
GJR-t	4.2878	8	390.1892	8	4.451	8	1.4186	8	1.5401	8	11.2284	8	9.5035	8	0.7	9.3015**
GJR-G	3.8177	2	365.7963	2	4.1768	2	1.2421	2	1.4585	3	10.8406	3	5.8995	2	0.79	14.1968**

Table 7: Out-of-sample predictability: long horizons

	7-step ahead volatility forecast															
Model	MSE1	Rank	MSE2	Rank	QLIKE	Rank	R2LOG	Rank	MAD2	Rank	MAD1	Rank	HMSE	Rank	SR	DA
GARCH-N	14.4585	5	1774.7723	5	7.8519	5	3.1028	6	3.3319	6	30.5326	6	39.1514	5	0.81	15.6381**
GARCH-t	18.3184	9	1929.3805	9	11.4239	9	5.1143	9	3.8675	9	32.7361	9	114.5564	9	0.75	13.1472**
GARCH-G	13.9793	1	1742.9475	1	7.5519	1	2.949	2	3.2864	2	30.3018	2	32.0057	1	0.84	17.5793**
EGARCH-N	14.3148	4	1770.9453	4	7.7573	4	3.0263	4	3.3007	4	30.392	4	38.0017	4	0.79	14.1349**
EGARCH-t	16.4193	7	1859.0518	7	9.4063	7	4.0483	7	3.6146	7	31.7511	7	68.3559	7	0.72	10.7186**
EGARCH-G	14.0432	2	1753.4072	3	7.5817	2	2.9391	1	3.2755	1	30.2646	1	33.6106	2	0.82	16.0067**
GJR-N	14.4688	6	1776.2451	6	7.8585	6	3.1007	5	3.3289	5	30.5209	5	39.4156	6	0.8	14.7728**
GJR-t	17.0239	8	1880.718	8	9.9887	8	4.387	8	3.7034	8	32.1067	8	79.6431	8	0.74	12.3028**
GJR-G	14.1045	3	1752.4542	2	7.6237	3	2.9828	3	3.2947	3	30.3486	3	33.6277	3	0.82	16.2786**

	14-step ahead volatility forecast															
Model	MSE1	Rank	MSE2	Rank	QLIKE	Rank	R2LOG	Rank	MAD2	Rank	MAD1	Rank	HMSE	Rank	SR	DA
GARCH-N	38.3654	5	6647.3113	5	13.9559	5	5.9558	5	5.7811	5	66.2684	5	164.7149	6	0.76	13.1921**
GARCH-t	50.0413	9	7159.2939	9	33.8145	9	11.699	9	6.7059	9	69.8064	9	1342.1712	9	0.66	8.4546**
GARCH-G	38.0254	2	6620.0351	2	13.68	2	5.8662	2	5.761	2	66.1657	2	152.3708	2	0.79	14.6833**
EGARCH-N	38.1732	3	6634.0569	4	13.7952	4	5.8956	3	5.7663	3	66.1971	3	157.7935	4	0.77	13.3317**
EGARCH-t	45.7174	7	6991.4793	7	23.0183	7	9.2003	7	6.3988	7	68.8086	7	524.6104	7	0.7	10.5994**
EGARCH-G	37.9858	1	6618.1606	1	13.6432	1	5.8489	1	5.756	1	66.1438	1	150.6241	1	0.77	13.2131**
GJR-N	38.4603	6	6649.7002	6	14.0261	6	5.9976	6	5.7919	6	66.3183	6	164.7069	5	0.78	13.8105**
GJR-t	47.2432	8	7053.0348	8	26.1577	8	10.0395	8	6.5128	8	69.1974	8	708.7607	8	0.67	9.5743**
GJR-G	38.1763	4	6627.2827	3	13.7936	3	5.9209	4	5.7747	4	66.2313	4	154.4992	3	0.78	13.7930**

	21-step ahead volatility forecast															
Model	MSE1	Rank	MSE2	Rank	QLIKE	Rank	R2LOG	Rank	MAD2	Rank	MAD1	Rank	HMSE	Rank	SR	DA
GARCH-N	64.8964	4	14445.7692	5	20.0519	5	8.0801	4	7.6592	4	101.9224	4	386.8934	5	0.72	11.4262**
GARCH-t	83.7149	9	15339.2163	9	69.2104	9	17.0065	9	8.789	9	106.0578	9	6589.1465	9	0.57	4.9508**
GARCH-G	64.7177	3	14427.951	3	19.8651	3	8.0422	3	7.6523	3	101.8866	3	371.912	3	0.77	12.6712**
EGARCH-N	64.7065	2	14427.9288	2	19.8462	2	8.0359	2	7.6507	2	101.88	2	369.6572	2	0.67	8.2322**
EGARCH-t	78.4454	7	15137.7856	7	44.1815	7	13.778	7	8.5072	7	105.2584	7	2153.9949	7	0.67	10.4329**
EGARCH-G	64.6305	1	14417.8386	1	19.7638	1	8.0261	1	7.6494	1	101.8713	1	361.066	1	0.7	8.4810**
GJR-N	65.0937	6	14453.9132	6	20.2138	6	8.1525	6	7.6754	6	101.9985	6	387.9121	6	0.71	10.6204**
GJR-t	80.5313	8	15221.9183	8	52.209	8	14.9836	8	8.6219	8	105.6	8	3209.7222	8	0.61	7.9106**
GJR-G	64.9183	5	14437.2193	4	20.0346	4	8.1128	5	7.6679	5	101.9602	5	374.5033	4	0.75	11.4336**

Table 8: Diebold-Mariano tests (horizon: one day, benchmark: GARCH-G)

Model	Diebold-Mariano							Modified Diebold-Mariano						
	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-4.70**	-2.95**	-4.84**	-6.93**	-6.97**	-8.52**	-2.89**	-4.70**	-2.95**	-4.83**	-6.92**	-6.96**	-8.51**	-2.89**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-t	-3.61**	-2.78**	-4.73**	-0.08	-3.35**	-2.97**	-3.82**	-3.61**	-2.77**	-4.73**	-0.08	-3.34**	-2.96**	-3.81**
p-values	0	0.01	0	0.93	0	0	0	0	0.01	0	0.93	0	0	0
EGARCH-N	-6.10**	-3.01**	-5.52**	-12.73**	-10.09**	-13.82**	-2.77**	-6.09**	-3.00**	-5.51**	-12.71**	-10.08**	-13.80**	-2.77**
p-values	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0.01
EGARCH-t	-5.75**	-3.23**	-6.81**	-4.78**	-5.57**	-6.20**	-3.31**	-5.74**	-3.22**	-6.80**	-4.78**	-5.57**	-6.19**	-3.31**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-G	-8.64**	-3.12**	-9.91**	-10.85**	-9.15**	-12.18**	-3.53**	-8.63**	-3.12**	-9.89**	-10.83**	-9.14**	-12.16**	-3.52**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-N	-5.29**	-2.97**	-5.13**	-10.60**	-8.34**	-11.11**	-2.83**	-5.29**	-2.96**	-5.13**	-10.58**	-8.32**	-11.09**	-2.83**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-t	-5.02**	-3.08**	-6.31**	-2.99**	-4.74**	-5.06**	-3.65**	-5.02**	-3.07**	-6.30**	-2.98**	-4.74**	-5.05**	-3.65**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-G	-7.67**	-3.14**	-9.74**	-9.16**	-7.78**	-10.75**	-4.23**	-7.66**	-3.14**	-9.73**	-9.15**	-7.77**	-10.74**	-4.22**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%, respectively.

Table 9: Diebold-Mariano tests (horizon: two days, benchmark: GARCH-G)

Model	Diebold-Mariano							Modified Diebold-Mariano						
	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-4.78**	-3.07**	-5.02**	-7.51**	-7.54**	-9.16**	-3.00**	-4.76**	-3.06**	-5.00**	-7.48**	-7.51**	-9.12**	-2.98**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-t	-4.20**	-3.04**	-6.57**	1	-3.20**	-2.61**	-5.37**	-4.19**	-3.02**	-6.55**	0.99	-3.19**	-2.60**	-5.35**
p-values	0	0	0	0.32	0	0.01	0	0	0	0	0.32	0	0.01	0
EGARCH-N	-4.65**	-2.89**	-4.80**	-9.41**	-7.47**	-9.42**	-2.95**	-4.63**	-2.88**	-4.78**	-9.37**	-7.44**	-9.38**	-2.94**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-t	-3.89**	-2.63**	-4.81**	-5.14**	-4.04**	-4.58**	-3.48**	-3.87**	-2.62**	-4.79**	-5.12**	-4.02**	-4.56**	-3.46**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0
EGARCH-G	-4.30**	-2.54*	-4.85**	-10.25**	-6.26**	-8.02**	-3.21**	-4.28**	-2.53*	-4.83**	-10.21**	-6.23**	-7.98**	-3.20**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0
GJR-N	-4.75**	-3.00**	-4.99**	-8.61**	-7.60**	-9.51**	-3.02**	-4.73**	-2.99**	-4.97**	-8.58**	-7.57**	-9.48**	-3.00**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-t	-4.04**	-2.77**	-5.50**	-3.43**	-4.10**	-4.49**	-4.13**	-4.02**	-2.76**	-5.47**	-3.41**	-4.09**	-4.48**	-4.11**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0
GJR-G	-4.75**	-2.77**	-5.84**	-10.77**	-6.92**	-9.09**	-3.95**	-4.73**	-2.75**	-5.82**	-10.73**	-6.89**	-9.06**	-3.93**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%, respectively.

Table 10: Diebold-Mariano tests (horizon: three days, benchmark: GARCH-G)

Model	Diebold-Mariano							Modified Diebold-Mariano						
	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-4.90**	-3.16**	-5.25**	-7.65**	-7.62**	-9.15**	-3.16**	-4.87**	-3.13**	-5.21**	-7.60**	-7.57**	-9.08**	-3.14**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-t	-7.75**	-4.06**	-11.00**	-10.65**	-11.18**	-12.96**	-6.96**	-7.70**	-4.03**	-10.93**	-10.58**	-11.10**	-12.88**	-6.91**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-N	-4.03**	-2.80**	-4.38**	-6.29**	-5.09**	-5.77**	-3.06**	-4.00**	-2.78**	-4.35**	-6.25**	-5.06**	-5.73**	-3.04**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0
EGARCH-t	-4.45**	-2.86**	-5.71**	-6.68**	-5.17**	-6.07**	-3.93**	-4.42**	-2.84**	-5.67**	-6.64**	-5.14**	-6.03**	-3.90**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-G	-2.98**	-2.30*	-3.29**	-4.59**	-2.66**	-2.62**	-2.99**	-2.96**	-2.29*	-3.27**	-4.56**	-2.65**	-2.60**	-2.97**
p-values	0	0.02	0	0	0.01	0.01	0	0	0.02	0	0	0.01	0.01	0
GJR-N	-4.58**	-3.02**	-4.99**	-7.21**	-6.57**	-7.75**	-3.21**	-4.55**	-3.00**	-4.96**	-7.16**	-6.53**	-7.70**	-3.19**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-t	-5.57**	-3.26**	-7.43**	-8.28**	-7.67**	-9.17**	-4.96**	-5.53**	-3.24**	-7.38**	-8.23**	-7.62**	-9.11**	-4.93**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-G	-3.93**	-2.66**	-4.77**	-6.38**	-4.53**	-5.08**	-3.87**	-3.91**	-2.64**	-4.74**	-6.34**	-4.50**	-5.04**	-3.85**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%, respectively.

Table 11: Diebold-Mariano tests (horizon: one week, benchmark: GARCH-G)

Model	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-4.83**	-3.24**	-5.24**	-7.08**	-6.68**	-7.66**	-3.50**	-4.75**	-3.18**	-5.15**	-6.96**	-6.56**	-7.53**	-3.44**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-t	-14.39**	-6.17**	-17.75**	-35.03**	-35.54**	-49.29**	-8.62**	-14.14**	-6.06**	-17.44**	-34.42**	-34.92**	-48.43**	-8.47**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-N	-3.07**	-2.55*	-3.38**	-3.16**	-2.20*	-1.97*	-2.93**	-3.02**	-2.51*	-3.32**	-3.10**	-2.16*	-1.93	-2.88**
p-values	0	0.01	0	0	0.03	0.05	0	0	0.01	0	0	0.03	0.05	0
EGARCH-t	-11.08**	-5.07**	-12.65**	-19.51**	-20.66**	-22.59**	-6.36**	-10.89**	-4.98**	-12.42**	-19.16**	-20.30**	-22.19**	-6.25**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-G	-1.13	-1.73	-1.09	0.76	1.49	2.41+	-2.11*	-1.11	-1.7	-1.07	0.74	1.47	2.36+	-2.08*
p-values	0.26	0.08	0.28	0.45	0.14	0.02	0.03	0.27	0.09	0.29	0.46	0.14	0.02	0.04
GJR-N	-4.67**	-3.19**	-5.10**	-6.51**	-5.89**	-6.54**	-3.55**	-4.59**	-3.13**	-5.02**	-6.39**	-5.79**	-6.43**	-3.49**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-t	-13.43**	-5.78**	-15.51**	-24.89**	-28.65**	-31.31**	-7.87**	-13.20**	-5.68**	-15.24**	-24.46**	-28.15**	-30.76**	-7.73**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-G	-3.80**	-2.78**	-4.44**	-4.29**	-3.28**	-3.20**	-3.90**	-3.73**	-2.73**	-4.36**	-4.21**	-3.23**	-3.15**	-3.83**
p-values	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%, respectively.

Table 12: Diebold-Mariano tests (horizon: one week, benchmark: EGARCH-G)

Model	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-6.78**	-4.43**	-6.60**	-9.52**	-10.42**	-11.20**	-3.76**	-6.66**	-4.35**	-6.49**	-9.35**	-10.24**	-11.00**	-3.70**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-t	-16.32**	-6.99**	-19.00**	-36.99**	-40.61**	-48.52**	-9.00**	-16.03**	-6.87**	-18.66**	-36.35**	-39.90**	-47.67**	-8.84**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-G	1.13	1.73	1.09	-0.76	-1.49	-2.41*	2.11+	1.11	1.7	1.07	-0.74	-1.47	-2.36*	2.08+
p-values	0.26	0.08	0.28	0.45	0.14	0.02	0.03	0.27	0.09	0.29	0.46	0.14	0.02	0.04
EGARCH-N	-4.44**	-3.29**	-4.48**	-5.51**	-5.46**	-5.71**	-3.17**	-4.36**	-3.23**	-4.40**	-5.42**	-5.36**	-5.61**	-3.12**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-t	-13.57**	-6.08**	-14.23**	-21.45**	-24.73**	-24.50**	-6.87**	-13.34**	-5.97**	-13.98**	-21.07**	-24.30**	-24.07**	-6.75**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-N	-7.11**	-4.58**	-6.88**	-10.00**	-11.21**	-12.03**	-3.98**	-6.99**	-4.50**	-6.76**	-9.82**	-11.02**	-11.82**	-3.91**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-t	-16.20**	-6.90**	-17.02**	-26.23**	-32.84**	-31.47**	-8.47**	-15.91**	-6.78**	-16.73**	-25.77**	-32.26**	-30.92**	-8.32**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-G	-2.36*	0.35	-3.30**	-7.13**	-7.09**	-8.61**	-0.05	-2.32*	0.34	-3.24**	-7.00**	-6.97**	-8.46**	-0.04
p-values	0.02	0.73	0	0	0	0	0.96	0.02	0.73	0	0	0	0	0.96

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%, respectively.

Table 13: Diebold-Mariano tests (horizon: two weeks, benchmark: EGARCH-G)

Model	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-2.72**	-2.70**	-3.83**	2.65++	-1.9	-1.46	-2.81**	-2.72**	-2.70**	-3.82**	2.65++	-1.89	-1.46	-2.80**
p-values	0.01	0.01	0	0.01	0.06	0.14	0.01	0.01	0.01	0	0.01	0.06	0.15	0.01
GARCH-t	4.16++	1.89	4.92++	4.68++	3.21++	3.89++	0.11	4.15++	1.88	4.91++	4.68++	3.21++	3.88++	0.11
p-values	0	0.06	0	0	0	0	0.91	0	0.06	0	0	0	0	0.91
GARCH-G	8.64++	3.12++	9.91++	10.85++	9.15++	12.18++	3.53++	8.63++	3.12++	9.89++	10.83++	9.14++	12.16++	3.52++
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EGARCH-N	-4.34**	-2.81**	-4.34**	-5.46**	-6.82**	-7.87**	-2.65**	-4.33**	-2.80**	-4.34**	-5.45**	-6.81**	-7.86**	-2.64**
p-values	0	0	0	0	0	0	0.01	0	0.01	0	0	0	0	0.01
EGARCH-t	-3.18**	-3.20**	-4.06**	-1.02	-2.76**	-2.58**	-3.00**	-3.18**	-3.20**	-4.05**	-1.02	-2.76**	-2.57*	-3.00**
p-values	0	0	0	0.31	0.01	0.01	0	0	0	0	0.31	0.01	0.01	0
GJR-N	-3.38**	-2.74**	-4.03**	0.05	-3.81**	-3.85**	-2.73**	-3.37**	-2.74**	-4.02**	0.05	-3.81**	-3.85**	-2.72**
p-values	0	0.01	0	0.96	0	0	0.01	0	0.01	0	0.96	0	0	0.01
GJR-t	-1.22	-2.79**	-2.07*	1.59	-0.96	-0.33	-3.40**	-1.21	-2.79**	-2.07*	1.59	-0.96	-0.33	-3.39**
p-values	0.22	0.01	0.04	0.11	0.34	0.74	0	0.23	0.01	0.04	0.11	0.34	0.74	0
GJR-G	9.75++	3.06++	9.51++	12.12++	10.63++	12.91++	2.51+	9.74++	3.06++	9.50++	12.10++	10.62++	12.89++	2.50+
p-values	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0.01

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%,
respectively.

Table 14: Diebold-Mariano tests (horizon: three weeks, benchmark: EGARCH-G)

Model	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE	MSE1	MSE2	QLIKE	R2LOG	MAD2	MAD1	HMSE
GARCH-N	-2.29*	-2.27*	-2.31*	-2.15*	-2.10*	-2.04*	-2.20*	-2.17*	-2.15*	-2.18*	-2.03*	-1.98*	-1.92	-2.08*
p-values	0.02	0.02	0.02	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.04	0.05	0.05	0.04
GARCH-t	-20.99**	-10.73**	-12.43**	-35.11**	-62.75**	-60.80**	-4.96**	-19.85**	-10.15**	-11.75**	-33.19**	-59.32**	-57.48**	-4.69**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GARCH-G	-1.26	-1.45	-1.35	-0.98	-0.96	-0.9	-1.58	-1.19	-1.37	-1.28	-0.93	-0.91	-0.85	-1.5
p-values	0.21	0.15	0.18	0.33	0.34	0.37	0.11	0.23	0.17	0.2	0.35	0.36	0.4	0.14
EGARCH-N	-1.35	-1.75	-1.47	-0.78	-0.69	-0.54	-1.9	-1.27	-1.65	-1.39	-0.74	-0.66	-0.51	-1.8
p-values	0.18	0.08	0.14	0.44	0.49	0.59	0.06	0.2	0.1	0.17	0.46	0.51	0.61	0.07
EGARCH-t	-25.86**	-11.84**	-19.65**	-40.23**	-57.52**	-52.29**	-8.33**	-24.45**	-11.20**	-18.58**	-38.03**	-54.38**	-49.43**	-7.88**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-N	-5.99**	-4.35**	-5.44**	-7.57**	-7.84**	-8.10**	-3.50**	-5.66**	-4.11**	-5.14**	-7.16**	-7.41**	-7.66**	-3.31**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-t	-23.54**	-11.30**	-16.48**	-36.38**	-57.61**	-52.58**	-7.03**	-22.25**	-10.68**	-15.58**	-34.39**	-54.46**	-49.70**	-6.65**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GJR-G	-8.40**	-5.94**	-7.29**	-9.86**	-10.17**	-10.24**	-4.01**	-7.94**	-5.62**	-6.89**	-9.33**	-9.62**	-9.68**	-3.79**
p-values	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Legend: * and ** indicate rejection of the null of equal predictive accuracy at 5% and 1%, respectively.

Table 15: Reality check and SPA Tests (horizon: one day)

Benchmark		MSE1	MSE2	QLIKE	R2LOG	MAD1	MAD2	HMSE
		Loss function						
GARCH-N	SPA_l^0	0	0.003	0	0.005	0	0	0.005
GARCH-N	SPA_c^0	0	0.003	0	0.006	0	0	0.005
GARCH-N	RC	0	0.003	0	0.013	0	0	0.005
GARCH-t	SPA_l^0	0.001	0.01	0	0.491	0.003	0.002	0.001
GARCH-t	SPA_c^0	0.088	0.156	0.234	0.774	0.033	0.018	0.419
GARCH-t	RC	0.09	0.158	0.235	0.789	0.033	0.018	0.426
GARCH-G	SPA_l^0	0.551	0.553	0.524	0.495	0.504	0.632	0.557
GARCH-G	SPA_c^0	1	0.977	1	0.949	1	1	0.689
GARCH-G	RC	1	0.994	1	0.953	1	1	1
EGARCH-N	SPA_l^0	0	0.007	0	0	0	0	0.007
EGARCH-N	SPA_c^0	0	0.007	0	0	0	0	0.007
EGARCH-N	RC	0	0.007	0	0	0	0	0.007
EGARCH-t	SPA_l^0	0	0.002	0	0	0	0	0.004
EGARCH-t	SPA_c^0	0	0.002	0	0	0	0	0.271
EGARCH-t	RC	0	0.002	0	0	0	0	0.272
EGARCH-G	SPA_l^0	0	0.006	0	0	0	0	0.003
EGARCH-G	SPA_c^0	0	0.05	0.044	0	0	0	0.413
EGARCH-G	RC	0	0.05	0.044	0	0	0	0.419
GJR-N	SPA_l^0	0	0.005	0	0	0	0	0.004
GJR-N	SPA_c^0	0	0.005	0	0	0	0	0.004
GJR-N	RC	0	0.005	0	0	0	0	0.004
GJR-t	SPA_l^0	0	0.003	0	0.008	0	0	0.002
GJR-t	SPA_c^0	0	0.003	0.005	0.008	0	0	0.295
GJR-t	RC	0	0.003	0.005	0.008	0	0	0.296
GJR-G	SPA_l^0	0	0.006	0	0.006	0	0	0
GJR-G	SPA_c^0	0.057	0.239	0.196	0.023	0.007	0.017	0.449
GJR-G	RC	0.057	0.242	0.197	0.023	0.007	0.017	0.461

Table 16: Reality check and SPA Tests (horizon: two days)

Benchmark		MSE1	MSE2	QLIKE	R2LOG	MAD1	MAD2	HMSE
					Loss function			
GARCH-N	SPA_l^0	0	0.005	0	0	0	0	0.004
GARCH-N	SPA_c^0	0	0.005	0	0	0	0	0.004
GARCH-N	RC	0	0.005	0	0	0	0	0.004
GARCH-t	SPA_l^0	0	0.005	0	0.524	0.004	0.001	0
GARCH-t	SPA_c^0	0	0.005	0	0.972	0.044	0.007	0.015
GARCH-t	RC	0	0.005	0	0.972	0.044	0.007	0.015
GARCH-G	SPA_l^0	0.547	0.537	0.534	0.164	0.509	0.553	0.534
GARCH-G	SPA_c^0	0.996	0.93	0.999	0.338	1	1	0.978
GARCH-G	RC	1	0.991	1	0.398	1	1	0.999
EGARCH-N	SPA_l^0	0	0.007	0	0	0	0	0.006
EGARCH-N	SPA_c^0	0	0.007	0	0	0	0	0.006
EGARCH-N	RC	0	0.007	0	0	0	0	0.006
EGARCH-t	SPA_l^0	0.001	0.016	0	0	0	0.001	0
EGARCH-t	SPA_c^0	0.001	0.016	0	0	0	0.001	0
EGARCH-t	RC	0.001	0.016	0	0	0	0.001	0
EGARCH-G	SPA_l^0	0	0.015	0	0	0	0	0.003
EGARCH-G	SPA_c^0	0	0.015	0.018	0	0	0	0.284
EGARCH-G	RC	0	0.015	0.019	0	0	0	0.285
GJR-N	SPA_l^0	0	0.006	0	0	0	0	0.003
GJR-N	SPA_c^0	0	0.006	0	0	0	0	0.003
GJR-N	RC	0	0.006	0	0	0	0	0.003
GJR-t	SPA_l^0	0	0.012	0	0	0	0	0
GJR-t	SPA_c^0	0	0.012	0	0	0	0	0
GJR-t	RC	0	0.012	0	0	0	0	0
GJR-G	SPA_l^0	0	0.012	0	0	0	0	0.001
GJR-G	SPA_c^0	0.025	0.012	0.122	0	0.002	0.003	0.159
GJR-G	RC	0.028	0.114	0.125	0	0.002	0.003	0.394

Table 17: Reality check and SPA Tests (horizon: three days)

Benchmark		Loss function						
		MSE1	MSE2	QLIKE	R2LOG	MAD1	MAD2	HMSE
GARCH-N	SPA_l^0	0	0.006	0	0	0	0	0.002
GARCH-N	SPA_c^0	0	0.006	0	0	0	0	0.002
GARCH-N	RC	0	0.006	0	0	0	0	0.002
GARCH-t	SPA_l^0	0	0.001	0	0	0	0	0
GARCH-t	SPA_c^0	0	0.001	0	0	0	0	0
GARCH-t	RC	0	0.001	0	0	0	0	0
GARCH-G	SPA_l^0	0.553	0.554	0.539	0.524	0.524	0.55	0.531
GARCH-G	SPA_c^0	0.98	0.554	0.993	0.999	0.997	0.995	0.973
GARCH-G	RC	1	0.988	1	1	1	1	1
EGARCH-N	SPA_l^0	0.001	0.009	0	0	0	0	0.004
EGARCH-N	SPA_c^0	0.001	0.009	0	0	0	0	0.004
EGARCH-N	RC	0.001	0.009	0	0	0	0	0.004
EGARCH-t	SPA_l^0	0.001	0.006	0	0	0	0	0.001
EGARCH-t	SPA_c^0	0.001	0.006	0	0	0	0	0.001
EGARCH-t	RC	0.001	0.006	0	0	0	0	0.001
EGARCH-G	SPA_l^0	0.006	0.023	0.004	0	0.009	0.009	0.007
EGARCH-G	SPA_c^0	0.006	0.023	0.024	0.019	0.164	0.051	0.194
EGARCH-G	RC	0.014	0.024	0.05	0.019	0.164	0.081	0.234
GJR-N	SPA_l^0	0	0.005	0	0	0	0	0.002
GJR-N	SPA_c^0	0	0.005	0	0	0	0	0.002
GJR-N	RC	0	0.005	0	0	0	0	0.002
GJR-t	SPA_l^0	0	0.004	0	0	0	0	0
GJR-t	SPA_c^0	0	0.004	0	0	0	0	0
GJR-t	RC	0	0.004	0	0	0	0	0
GJR-G	SPA_l^0	0.001	0.011	0	0	0	0.001	0.001
GJR-G	SPA_c^0	0.023	0.011	0.071	0.035	0.056	0.034	0.275
GJR-G	RC	0.077	0.164	0.138	0.052	0.113	0.092	0.348

Table 18: Reality check and SPA Tests (horizon: one week)

Benchmark		MSE1	MSE2	QLIKE	R2LOG	MAD1	MAD2	HMSE
					Loss function			
GARCH-N	SPA_l^0	0	0.006	0	0	0	0	0.006
GARCH-N	SPA_c^0	0	0.006	0	0	0	0	0.006
GARCH-N	RC	0	0.006	0	0	0	0	0.006
GARCH-t	SPA_l^0	0	0.007	0	0.459	0	0	0.002
GARCH-t	SPA_c^0	0.075	0.137	0.218	0.768	0.014	0.006	0.454
GARCH-t	RC	0.076	0.138	0.218	0.781	0.014	0.006	0.464
GARCH-G	SPA_l^0	0.55	0.562	0.524	0.512	0.507	0.594	0.549
GARCH-G	SPA_c^0	1	0.975	1	0.931	1	1	0.686
GARCH-G	RC	1	0.998	1	0.935	1	1	1
EGARCH-N	SPA_l^0	0	0.005	0	0	0	0	0.008
EGARCH-N	SPA_c^0	0	0.005	0	0	0	0	0.008
EGARCH-N	RC	0	0.005	0	0	0	0	0.008
EGARCH-t	SPA_l^0	0	0.001	0	0	0	0	0.001
EGARCH-t	SPA_c^0	0	0.001	0	0	0	0	0.266
EGARCH-t	RC	0	0.001	0	0	0	0	0.266
EGARCH-G	SPA_l^0	0	0.002	0	0	0	0	0.001
EGARCH-G	SPA_c^0	0	0.04	0.05	0	0	0	0.433
EGARCH-G	RC	0	0.04	0.05	0	0	0	0.44
GJR-N	SPA_l^0	0	0.007	0	0	0	0	0.005
GJR-N	SPA_c^0	0	0.007	0	0	0	0	0.005
GJR-N	RC	0	0.007	0	0	0	0	0.005
GJR-t	SPA_l^0	0	0.003	0	0	0	0	0.001
GJR-t	SPA_c^0	0	0.003	0.002	0	0	0	0.299
GJR-t	RC	0	0.003	0.002	0	0	0	0.299
GJR-G	SPA_l^0	0	0.002	0	0.001	0	0	0
GJR-G	SPA_c^0	0.045	0.233	0.2	0.002	0	0.001	0.446
GJR-G	RC	0.045	0.236	0.2	0.002	0	0.001	0.457

Table 19: Reality check and SPA Tests (horizon: two weeks)

Benchmark		MSE1	MSE2	QLIKE	R2LOG	MAD1	MAD2	HMSE
				Loss function				
GARCH-N	SPA_l^0	0	0.003	0	0	0	0	0.004
GARCH-N	SPA_c^0	0	0.003	0	0	0	0	0.004
GARCH-N	RC	0	0.003	0	0	0	0	0.004
GARCH-t	SPA_l^0	0	0.002	0	0.501	0.006	0	0
GARCH-t	SPA_c^0	0	0.002	0	0.981	0.033	0.003	0.022
GARCH-t	RC	0	0.002	0	0.981	0.033	0.003	0.022
GARCH-G	SPA_l^0	0.532	0.519	0.521	0.16	0.508	0.55	0.521
GARCH-G	SPA_c^0	1	0.961	1	0.315	1	1	0.986
GARCH-G	RC	1	0.996	1	0.376	1	1	1
EGARCH-N	SPA_l^0	0	0.003	0	0	0	0	0.004
EGARCH-N	SPA_c^0	0	0.003	0	0	0	0	0.004
EGARCH-N	RC	0	0.003	0	0	0	0	0.004
EGARCH-t	SPA_l^0	0	0.002	0	0	0	0	0
EGARCH-t	SPA_c^0	0	0.002	0	0	0	0	0.001
EGARCH-t	RC	0	0.002	0	0	0	0	0.001
EGARCH-G	SPA_l^0	0	0.004	0	0	0	0	0.001
EGARCH-G	SPA_c^0	0	0.004	0.023	0	0	0	0.297
EGARCH-G	RC	0	0.004	0.023	0	0	0	0.301
GJR-N	SPA_l^0	0	0.003	0	0	0	0	0.003
GJR-N	SPA_c^0	0	0.003	0	0	0	0	0.003
GJR-N	RC	0	0.003	0	0	0	0	0.003
GJR-t	SPA_l^0	0	0.002	0	0	0	0	0
GJR-t	SPA_c^0	0	0.002	0	0	0	0	0
GJR-t	RC	0	0.002	0	0	0	0	0
GJR-G	SPA_l^0	0	0.002	0	0	0	0	0
GJR-G	SPA_c^0	0.011	0.002	0.121	0	0.001	0.001	0.128
GJR-G	RC	0.012	0.065	0.122	0	0.001	0.001	0.382

Table 20: Reality check and SPA Tests (horizon: three weeks)

Benchmark		MSE1	MSE2	QLIKE	R2LOG	MAD1	MAD2	HMSE
					Loss function			
GARCH-N	SPA_l^0	0	0.001	0	0	0	0	0.002
GARCH-N	SPA_c^0	0	0.001	0	0	0	0	0.002
GARCH-N	RC	0	0.001	0	0	0	0	0.002
GARCH-t	SPA_l^0	0	0	0	0	0	0	0
GARCH-t	SPA_c^0	0	0	0	0	0	0	0
GARCH-t	RC	0	0	0	0	0	0	0
GARCH-G	SPA_l^0	0.527	0.528	0.512	0.513	0.502	0.536	0.517
GARCH-G	SPA_c^0	0.995	0.528	0.998	1	1	0.999	0.985
GARCH-G	RC	1	0.998	1	1	1	1	1
EGARCH-N	SPA_l^0	0	0.002	0	0	0	0	0.001
EGARCH-N	SPA_c^0	0	0.002	0	0	0	0	0.001
EGARCH-N	RC	0	0.002	0	0	0	0	0.001
EGARCH-t	SPA_l^0	0	0.001	0	0	0	0	0
EGARCH-t	SPA_c^0	0	0.001	0	0	0	0	0
EGARCH-t	RC	0	0.001	0	0	0	0	0
EGARCH-G	SPA_l^0	0.001	0.009	0.001	0	0.001	0	0.003
EGARCH-G	SPA_c^0	0.001	0.009	0.024	0.021	0.119	0.036	0.195
EGARCH-G	RC	0.002	0.009	0.04	0.021	0.119	0.047	0.223
GJR-N	SPA_l^0	0	0.001	0	0	0	0	0.001
GJR-N	SPA_c^0	0	0.001	0	0	0	0	0.001
GJR-N	RC	0	0.001	0	0	0	0	0.001
GJR-t	SPA_l^0	0	0.001	0	0	0	0	0
GJR-t	SPA_c^0	0	0.001	0	0	0	0	0
GJR-t	RC	0	0.001	0	0	0	0	0
GJR-G	SPA_l^0	0	0.005	0	0	0	0	0
GJR-G	SPA_c^0	0.007	0.005	0.067	0.038	0.06	0.026	0.3
GJR-G	RC	0.028	0.076	0.114	0.044	0.096	0.049	0.361

Table 21: Out-of-sample evaluation of risk management (short horizons)

1-step ahead

Model	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF
			95% VaR							99% VaR				
GARCH-N	4	5.676	0.683	4.648*	5.331	0.3206	0.288	21	2.162	7.577*	5.166*	12.743*	0.1844	0.1533
GARCH-t	4	**2.568**	11.131*	2.796	13.927*	0.194	0.1343	60	**0.541**	1.894	**0.558**	2.452	0.125	0.0098
GARCH-G	4	4.595	0.263	3.69	**3.953**	0.284	0.2453	60	0.811	**0.286**	0.718	**1.004**	0.1473	0.0771
EGARCH-N	4	4.865	**0.029**	4.971*	4.999	0.2972	0.2607	21	2.162	7.577*	5.166*	12.743*	0.1707	0.1306
EGARCH-t	21	**2.568**	11.131*	**2.45**	13.580*	0.173	0.1107	60	**0.541**	1.894	**0.558**	2.452	0.1218	**0.002**
EGARCH-G	4	4.459	0.472	4.971*	5.442	0.2699	0.2296	60	0.676	0.887	**0.558**	1.445	0.1432	0.0694
GJR-N	4	5.405	0.25	4.648*	4.898	0.3062	0.2714	21	2.162	7.577*	5.166*	12.743*	0.1761	0.1398
GJR-t	4	**2.568**	11.131*	2.796	13.927*	**0.1818**	**0.1212**	60	**0.541**	1.894	**0.558**	2.452	**0.1213**	0.0045
GJR-G	4	4.73	0.116	4.851*	4.967	0.2776	0.2385	60	0.676	0.887	**0.558**	1.445	0.1424	0.0697

2-step ahead

Model	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF
			95% VaR							99% VaR				
GARCH-N	3	5.541	0.44	8.241*	8.681*	0.2981	0.2637	20	2.162	7.577*	5.166*	12.743*	0.1674	0.1288
GARCH-t	3	2.973	7.443*	4.357*	11.800*	0.2182	0.1646	59	**0.541**	1.894	**0.558**	2.452	0.1279	0.0225
GARCH-G	3	4.595	0.263	3.69	**3.953**	0.2871	0.2488	59	**0.541**	1.894	**0.558**	2.452	0.1454	0.0736
EGARCH-N	3	5	**0**	9.439*	9.439*	0.284	0.2467	20	2.162	7.577*	5.166*	12.743*	0.1606	0.1162
EGARCH-t	3	2.973	7.443*	4.357*	11.800*	**0.183**	**0.1239**	59	**0.541**	1.894	**0.558**	2.452	**0.1197**	**0.0037**
EGARCH-G	3	4.324	0.744	5.296*	6.040*	0.2777	0.2379	59	0.676	0.887	**0.558**	1.445	0.1418	0.0672
GJR-N	3	5	**0**	9.439*	9.439*	0.29	0.254	20	2.162	7.577*	5.166*	12.743*	0.1635	0.1217
GJR-t	3	**2.838**	8.573*	4.585*	13.157*	0.1963	0.1396	59	**0.541**	1.894	**0.558**	2.452	0.1204	0.0087
GJR-G	3	4.459	0.472	**3.662**	4.134	0.2829	0.2441	59	0.811	**0.286**	0.718	**1.004**	0.1399	0.0661

3-step ahead

Model	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF
			95% VaR							99% VaR				
GARCH-N	2	5.135	0.028	9.439*	9.467*	0.304	0.2698	19	2.297	9.205*	4.853*	14.058*	0.1735	0.1369
GARCH-t	2	3.649	3.127	**3.718**	**6.845***	0.228	0.1793	58	**0.541**	**1.894**	**0.558**	**2.452**	0.1258	0.0273
GARCH-G	2	4.595	0.263	7.432*	7.695*	0.2745	0.2354	58	**0.541**	**1.894**	**0.558**	**2.452**	0.1394	0.0647
EGARCH-N	2	4.865	0.029	10.438*	10.466*	0.2901	0.2532	19	1.757	3.493	2.687	6.180*	0.1656	0.1225
EGARCH-t	2	**2.973**	7.443*	4.357*	11.800*	**0.1881**	**0.1314**	58	**0.541**	**1.894**	**0.558**	**2.452**	**0.1177**	0.0051
EGARCH-G	2	4.189	1.081	7.800*	8.882*	0.2674	0.227	58	**0.541**	**1.894**	**0.558**	**2.452**	0.1358	0.0583
GJR-N	2	5	**0**	9.917*	9.917*	0.2963	0.2608	19	2.162	7.577*	5.166*	12.743*	0.1689	0.129
GJR-t	2	3.108	6.407*	4.166*	10.573*	0.2027	0.1491	58	**0.541**	**1.894**	**0.558**	**2.452**	0.1185	**0.0113**
GJR-G	2	4.595	0.263	7.102*	7.365*	0.2706	0.2313	58	**0.541**	**1.894**	**0.558**	**2.452**	0.1353	0.059

Legend: * indicates significance at the 5% level.

Table 22: Out-of-sample evaluation of risk management (long horizons)

7-step ahead

	95% VaR						99% VaR					
Model	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF
GARCH-N	15	4.459	0.472	9.917*	10.388*	0.2874	0.2517	54	1.757	3.493	2.687	6.180*	0.162	0.1185
GARCH-t	4	5.541	0.44	**7.913***	**8.353***	0.3234	0.2947	54	1.081	**0.048**	1.064	**1.112**	0.1418	0.075
GARCH-G	15	4.324	0.744	10.438*	11.182*	0.2745	0.236	54	0.811	0.286	0.887	1.173	0.1424	0.0683
EGARCH-N	15	4.324	0.744	10.438*	11.182*	0.2803	0.243	54	1.486	1.539	3.035	4.574	0.158	0.1108
EGARCH-t	15	**4.189**	1.081	11.621*	12.703*	**0.2368**	**0.1922**	54	**0.541**	1.894	**0.558**	2.452	**0.1193**	**0.023**
EGARCH-G	15	4.324	0.744	10.438*	11.182*	0.2696	0.2304	54	**0.541**	1.894	**0.558**	2.452	0.1361	0.0582
GJR-N	15	4.324	0.744	10.438*	11.182*	0.2852	0.2491	54	1.622	2.431	2.838	5.27	0.1604	0.1156
GJR-t	15	4.595	**0.263**	9.917*	10.180*	0.2639	0.2248	54	0.676	0.887	0.718	1.606	0.1245	0.038
GJR-G	15	**4.189**	1.081	11.005*	12.086*	0.2723	0.2337	54	0.811	0.286	0.887	1.173	0.1377	0.0621

14-step ahead

	95% VaR						99% VaR					
Model	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF
GARCH-N	8	4.459	0.472	9.917*	10.388*	0.2893	0.2454	44	1.622	2.431	2.838	5.27	0.1649	0.1217
GARCH-t	8	8.919	19.606*	7.657*	27.263*	0.4849	0.4675	8	2.162	7.577*	5.166*	12.743*	0.2054	0.1943
GARCH-G	8	**4.324**	0.744	7.800*	**8.544***	**0.2719**	0.226	47	**0.676**	0.887	**0.718**	1.606	0.1309	0.0513
EGARCH-N	8	4.459	0.472	9.917*	10.388*	0.2845	0.24	47	1.486	1.539	3.035	4.574	0.162	0.1164
EGARCH-t	8	6.081	1.708	7.913*	9.621*	0.3332	0.2977	47	1.081	**0.048**	1.25	**1.298**	0.1372	0.0713
EGARCH-G	8	4.324	0.744	7.800*	**8.544***	0.2698	**0.2236**	47	**0.676**	0.887	**0.718**	1.606	0.1292	**0.0483**
GJR-N	8	4.595	**0.263**	9.917*	10.180*	0.2893	0.2456	44	1.622	2.431	2.838	5.27	0.164	0.1207
GJR-t	8	7.162	6.461*	**6.372***	12.833*	0.3814	0.3524	8	1.622	2.431	2.838	5.27	0.1549	0.1067
GJR-G	8	4.459	0.472	10.438*	10.909*	0.2731	0.2276	47	**0.676**	0.887	**0.718**	1.606	**0.1308**	0.0518

21-step ahead

	95% VaR						99% VaR					
Model	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF	TUFF	PF(%)	LRPF	LRind	LRcc	FLF	RLF
GARCH-N	1	4.459	**0.472**	9.439*	9.911*	0.2916	0.2497	40	1.486	1.539	2.838	4.377	0.1687	0.109
GARCH-t	1	10.946	41.745*	8.873*	50.618*	0.6399	0.6425	1	3.108	21.299*	9.155*	30.454*	0.3055	0.3056
GARCH-G	1	**4.324**	0.744	7.102*	7.846*	0.2845	0.2413	40	0.946	0.022	1.064	**1.086**	0.1411	0.0569
EGARCH-N	1	**4.324**	0.744	7.102*	7.846*	0.2868	0.2442	40	1.351	0.831	3.035	3.867	0.1651	0.1038
EGARCH-t	1	7.838	10.779*	**6.222***	17.001*	0.4506	0.4322	1	2.027	6.076*	2.496	8.573*	0.19	0.1421
EGARCH-G	1	4.189	1.081	7.432*	8.514*	0.2783	**0.2346**	40	**0.811**	0.286	**0.887**	1.173	**0.1383**	0.0529
GJR-N	1	4.459	**0.472**	7.102*	**7.573***	0.2915	0.2499	40	1.486	1.539	2.838	4.377	0.1672	0.1075
GJR-t	1	8.649	17.190*	6.919*	24.109*	0.5174	0.507	1	2.703	14.788*	10.961*	25.749*	0.226	0.1955
GJR-G	1	**4.324**	0.744	7.102*	7.846*	**0.2853**	0.2425	40	0.946	**0.022**	1.064	**1.086**	0.1403	**0.0566**

Legend: * indicates significance at the 5% level.