SCHOOL OF ECONOMICS AND MANAGEMENT
DEPARTMENT OF ECONOMICS

Master Essay II — NEKP01

Forecasting Swedish Stock Market Volatility and Value-at-Risk: A Comparison of EWMA and GARCH Models

Carl Nilsson
carl.nilsson.949@student.lu.se

supervised by
Birger Nilsson

August 27, 2017

Abstract

In this study we compare different volatility models on their ability to forecast one day ahead volatility and value-at-risk (VaR). We compare five different GARCH specifications: GARCH, IGARCH, GJR-GARCH, EGARCH and APARCH, as well as EWMA, each paired with six different conditional distributions. These models are used to forecast volatility and VaR one day ahead using daily return data from the Swedish stock market index OMXS30. The forecasts are then compared using the model confidence set procedure of Peter Reinhard Hansen, Asger Lunde, and James M Nason (2011). "The model confidence set." In: *Econometrica* 79.2, pp. 453–497.

We find the APARCH models best for forecasting volatility, while for forecasting VaR the best models are either APARCH, GJR-GARCH or EGARCH—depending on which level of VaR we use—paired with conditional distributions that take skewness and excess kurtosis into account. EWMA, GARCH and IGARCH specifications cannot be recommended either for forecasting volatility or for forecasting VaR.

Keywords: volatility forecasting, VaR, GARCH, model confidence set

Contents

List of Figures

List of Tables

1 Introduction

ARCH (autoregressive conditional heteroskedasticity) models the conditional volatility of a time series that is non-constant (conditionally heteroskedastic), this is done by modeling the squared residuals as an autoregressive (AR) process, hence the name.

Ever since the publication of the ARCH model in the Nobel Prize winning paper by Engle (1982) and the subsequent GARCH (generalized ARCH) model by Bollerslev (1986), there have been an explosion of different GARCH models.[1] Reviews can be found in Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994) and Teräsvirta (2009); Bollerslev (2008) describes itself as an "easy-to-use encyclopedic type reference guide to the long list of ARCH acronyms" and tries to cover all the published GARCH models up to that date, and can be very convenient when you feel lost in the GARCH acronym jungle.

The primary purpose of this thesis is to investigate which of the many GARCH models[2] (or rather, which of the ones—subjectively and quite arbitrarily—chosen by the author) that performs the best for forecasting (i) one day ahead volatility and (ii) one day ahead value-at-risk (VaR), using out of sample daily data of the Swedish stock market price index OMX Stockholm 30 (OMXS30). The newly developed model confidence set (MCS) procedure of Hansen, Lunde, and Nason (2011) is used for selecting the "best" models.

A secondary objective is also to find out if there are differences in performance in forecasting volatility versus VaR, if different models are "best" (as measured by lowest statistical loss or inclusion and rank by MCS) at one or the other, or if the best models excel at both tasks. And also, if the VaR^α model ranks are consistent for different values of α.

Finally, we also want to find out if the model specification or conditional distribution used is the more important for each of the tasks. How helpful a distribution that allows for skewness and excess kurtosis is for forecasting conditional volatility remains to be seen but our guess is that it will probably be more important for forecasting value-at-risk, which is a quantile function, and where the negative tail of the distribution is the only part of interest.

The VaR forecasts are also tested using Christoffersen's conditional coverage test (Christoffersen 1998) which tests the joint hypothesis of correct number of exceedances and that the exceedances are independent.

There have been many previous studies comparing volatility forecasting models, see, e.g., Andersen, Bollerslev, et al. (2006) for a theoretical overview and Poon

[1] For example, Hansen and Lunde (2005) compares 330 different GARCH specifications.

[2] The exponential weighted moving average is also included as a benchmark model to compare against.

and Granger (2003, 2005) for extensive surveys on the previous research. However, Patton and Sheppard (2009) finds "that the use of loss functions that are 'non-robust', in the sense of Patton (2006), can yield perverse rankings of forecasts, even when accurate volatility proxies are employed." Brownlees, Engle, and Kelly (2011b) adds that "the conflicting evidence reported by some previously published empirical studies on volatility forecasting is due to the use of non robust losses, and this calls for a reassessment of previous findings in this field." This thesis only uses those statistical loss functions that are "robust" as defined by Patton (2006, 2011)[3], namely quasi-likelihood (QL) and mean squared error (MSE), when examining the volatility forecasts.

The contributions of this thesis are threefold: firstly, we use the same models and data to to estimate both the conditional volatility and value-at-risk forecasts, making the results comparable; secondly, we use the newly developed model confidence set procedure of Hansen, Lunde, and Nason (2011) to select the models from the above mentioned forecasts; and thirdly, we use data from Sweden which—to our knowledge—has not been used in a similar study before.

Our findings are that APARCH is best for forecasting one-day-ahead volatility. The model specification is important, for MSE losses only APARCH-t and APARCH-skew t remains in the MCS, while for QL losses, all APARCH, 4 EGARCH and 4 GJR-models remain, with all the EWMA, IGARCH and GARCH models eliminated. The conditional distribution does not seem to make that much of a contribution. For value-at-risk, we find that conditional distributions are much more important, but the model specification remains so as well. EWMA, IGARCH and GARCH models are all eliminated and the best performing models differ depending on if we forecast $\text{VaR}^{1\%}$, $\text{VaR}^{5\%}$ or $\text{VaR}^{10\%}$. What all the best performing models have in common is that they are conditionally distributed with distributions are capable of capturing skewness and excess kurtosis.

The remainder of the thesis is organized as follows. In Section 2 we describe the conditional volatility models and conditional distributions used to make the out-of-sample forecasts. Section 3 is the main section where we describe the backtests and the results and compare them to some previous studies. In Section 4 we summarize the thesis and add some suggestions for further research. We have also included three appendices, Appendix A with plots of one day ahead forecasts of $\hat{\sigma}_t$, Appendix B with plots of the estimated parameters, and Appendix C with plots of the estimated VaR together with the returns.

[3] Patton (2006) is a longer, working paper version of Patton (2011).

2 Volatility models and distributions

We have a series of daily closing prices $\{P_t\}$ and calculate the percent log-returns $\{r_t\}$, defined as

$$r_t = 100\,(\ln P_t - \ln P_{t-1})\,. \tag{2.1}$$

The returns are modeled as[4]

$$r_t = \mu + \sigma_t z_t, \qquad z_t|\mathscr{F}_{t-1} \overset{\text{iid}}{\sim} \mathcal{D}(0,1,\cdot) \tag{2.2}$$

where $\mathscr{F}_{t-1}$ is the information set with all available information at time $t-1$. We let $\varepsilon_t = \sigma_t z_t = r_t - \mu$ denote the error. The standardized error is denoted $z_t = \varepsilon_t/\sigma_t$, which is independent and identically distributed (iid) with zero mean and unit variance and distribution $\mathcal{D}$ (see Section 2.2); $\cdot$ denotes any potential additional parameters (e.g., for skewness and/or excess kurtosis). The conditional variance $\sigma_t^2 \equiv \text{Var}\,(r_t|\mathscr{F}_{t-1})$ is modeled according to the specific volatility model described below and summarized in Table 2.1.

2.1 Conditional volatility models

2.1.1 EWMA

The first and simplest volatility model we use is the exponential weighted moving average (EWMA), defined as

$$\sigma_t^2 = (1-\lambda)\sum_{j=1}^{t-1}\lambda^{j-1}(r_{t-j} - \hat{\mu}_t)^2, \qquad \hat{\mu}_t = \frac{1}{t-1}\sum_{j=1}^{t-1} r_{t-j} \tag{2.3}$$

The parameter λ could be estimated, but we fix it to $\lambda = 0.94$ as suggested by RiskMetrics (J.P. Morgan and Reuters 1996), and then there are no parameters to estimate (except for potential distributional parameters). EWMA is equivalent to IGARCH (to be defined in Section 2.1.3) with $\omega = 0$ and $\beta = \lambda$ (see e.g., Dowd 2005, p. 135), i.e.,

$$\sigma_t^2 = (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2.$$

2.1.2 GARCH

The second model that we estimate is the standard GARCH (generalized autoregressive conditional heteroskedasticity) of Bollerslev (1986):

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2.4}$$

To ensure that σ_t^2 stays positive we have the following parameter restrictions $\omega > 0$, $\alpha, \beta \geq 0$, and to ensure stationarity: $\alpha + \beta < 1$.

[4] We have also tried an AR(1) specification, $r_t = \mu + \varphi r_{t-1} + \varepsilon_t$, which made no noticeable difference for the volatility estimates.

2.1.3 IGARCH

The third model is the integrated GARCH (IGARCH) of Engle and Bollerslev (1986), which is just the standard GARCH (2.4) with $\alpha + \beta = 1$, or

$$\sigma_t^2 = \omega + (1-\beta)\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{2.5}$$

2.1.4 GJR-GARCH

The fourth model is an asymmetric model, GJR-GARCH of Glosten, Jagannathan, and Runkle (1993) also called threshold GARCH (TARCH or TGARCH):[5]

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma \mathbb{1}_{\{\varepsilon_{t-1}\leq 0\}}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2.6}$$

where $\mathbb{1}_{\{\varepsilon_{t-1}\leq 0\}}$ is the indicator function which takes the value 1 when $\varepsilon_{t-1} \leq 0$ and 0 otherwise. The parameter γ captures the asymmetry, or the extent to which volatility increases more following negative shocks than for positive shocks (as long as $\gamma > 0$, which tends to be the case for equity and equity index returns).[6]

2.1.5 EGARCH

The fifth model is the exponential GARCH (EGARCH) of Nelson (1991), which models the log of conditional variance:

$$\ln\sigma_t^2 = \omega + \alpha z_{t-1} + \gamma\left(|z_{t-1}| - \mathrm{E}\left[|z_{t-1}|\right]\right) + \beta\ln\sigma_{t-1}^2. \tag{2.7}$$

The natural logarithm is used to prevent σ_t^2 from becoming negative, consequently EGARCH does not need to have any parameter restrictions, which adds flexibility to the model (Nelson 1991). Here α captures the sign effect and γ the size effect. z_{t-1} is the standardized residual $z_{t-1} = \varepsilon_{t-1}/\sigma_{t-1}$. The expectation $\mathrm{E}\left[|z_{t-1}|\right]$ depends on the distribution of z_t, e.g., when $z_t \sim \mathcal{N}(0,1)$, $\mathrm{E}\left[|z_{t-1}|\right] = \sqrt{2/\pi}$ and when $z_t \sim t_\nu$, $\mathrm{E}\left[|z_{t-1}|\right] = \frac{2\sqrt{\nu-2}\Gamma[(\nu+1)/2]}{\sqrt{\pi}(\nu-1)\Gamma[\nu/2]}$, where $\Gamma(\cdot)$ is the gamma function (Taylor 2005, Eqs. 10.2 & 10.4).

2.1.6 APARCH

The sixth and final model is the asymmetric power ARCH (APARCH) of Ding, Granger, and Engle (1993):

$$\sigma_t^\delta = \omega + \alpha\left(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1}\right)^\delta + \beta\sigma_{t-1}^\delta \tag{2.8}$$

[5] There is however another model also called threshold GARCH, namely the one by Zakoian (1994), so we will use the name GJR-GARCH to avoid confusion.

[6] The asymmetry is also called the leverage effect since Black (1976) believed it existed due to financial leverage, which has since then been proven to be incorrect. (Taylor 2005, pp. 241–242)

The parameter $-1 < \gamma < 1$ captures asymmetries, while the parameter $\delta \in \mathbb{R}^+$ helps to capture volatility dynamics more flexibly than for other specifications.[7] APARCH nests several other GARCH models (Ding, Granger, and Engle 1993, Appendix A; Hentschel 1995; Bollerslev 2008, p. 3), e.g., it reduces to GJR-GARCH when $\delta = 2$ and $0 < \gamma < 1$ (see Ding, Granger, and Engle 1993, Appendix A for the calculations), and to GARCH when $\delta = 2$ and $\gamma = 0$.

A summary of all the models is found in Table 2.1.

Table 2.1 The volatility models

Name	Model
EWMA	$\sigma_t^2 = (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2$, $\lambda = 0.94$
GARCH	$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
IGARCH	$\sigma_t^2 = \omega + (1-\beta)\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
GJR-GARCH	$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\mathbb{1}_{\{\varepsilon_{t-1}\leq 0\}}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
EGARCH	$\ln\sigma_t^2 = \omega + \alpha z_{t-1} + \gamma\left(\lvert z_{t-1}\rvert - \mathrm{E}\left[\lvert z_{t-1}\rvert\right]\right) + \beta\ln\sigma_{t-1}^2$ where $z_t = \varepsilon_t/\sigma_t$
APARCH	$\sigma_t^\delta = \omega + \alpha\left(\lvert\varepsilon_{t-1}\rvert - \gamma\varepsilon_{t-1}\right)^\delta + \beta\sigma_{t-1}^\delta$

2.2 Conditional distributions

The volatility models are all estimated with each of the following conditional distributions for $z_t|\mathscr{F}_{t-1}$.

2.2.1 Normal (N)

Completely described by its mean and variance, here scaled to a standard normal, with zero mean and unit variance. No skewness or excess kurtosis. The density is given by:

$$f_{\mathcal{N}}(z_t|\mathscr{F}_{t-1}) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_t^2\right)$$

2.2.2 Student's t (t)

As used for GARCH models by Bollerslev (1987). Described completely by its shape parameter ν which allows excess kurtosis (as ν goes to ∞, Student's t distribution converges to the normal distribution and the excess kurtosis goes to

[7] One stylized fact for financial returns, the Taylor effect—first mentioned in Taylor (1986)—says that for most series the autocorrelation is higher for $|r_t|$ than for r_t^2 and Ding, Granger, and Engle (1993) and Granger and Ding (1995) shows that it peaks around $k = 1$ for $|r_t|^k$. For the OMXS30 daily return data used in this thesis the autocorrelation peaks at $k = 1.25$ with a value of 0.214 and the autocorrelation of $|r_t|$ is greater than that of r_t^2, with values of 0.213 and 0.186, respectively.

0). No skewness. The density is given by:

$$f_t(z_t|\mathscr{F}_{t-1}) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{z_t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where $\Gamma(\cdot)$ is the gamma function.

2.2.3 Skew Normal (skew N)

The skew normal is the normal distribution scaled with a skewness parameter $\xi \in (0, \infty)$, as formulated by Fernandez and Steel (1998). The density is given by:

$$f_{S\mathcal{N}}(z_t|\mathscr{F}_{t-1}) = \frac{2}{\xi + \frac{1}{\xi}}\left\{f_{\mathcal{N}}\left(\frac{z_t}{\xi}\right)\mathbb{1}_{\{z_t \geq 0\}} + f_{\mathcal{N}}(\xi z_t)\mathbb{1}_{\{z_t<0\}}\right\} \tag{2.9}$$

where $f_{\mathcal{N}}(\cdot)$ is the density of the normal distribution, and where $\mathbb{1}_{\{x\}}$ denotes the indicator function which takes the value 1 when the condition x is fulfilled and 0 otherwise.

2.2.4 Skew Student's t (skew t)

Skew Student's t is the Student's t distribution scaled with a skewness parameter $\xi \in (0, \infty)$, also by Fernandez and Steel (1998). The density is the same as $f_{S\mathcal{N}}(z_t|\mathscr{F}_{t-1})$ in (2.9) but with $f_{\mathcal{N}}(\cdot)$ replaced by $f_t(\cdot)$—the density of Student's t distribution.

2.2.5 Normal Inverse Gaussian (NIG)

Allows for skewness and excess kurtosis. See Barndorff-Nielsen (1997). The density is given by:

$$f_{NIG}(z_t|\mathscr{F}_{t-1}) = \frac{\alpha\delta}{\pi}\exp\left(\delta\sqrt{\alpha^2-\beta^2}+\beta(z_t-\mu)\right)\frac{K_1\left(\alpha\sqrt{\delta^2+(z_t-\mu)^2}\right)}{\sqrt{\delta^2+(z_t-\mu)^2}}$$

$K_1(\cdot)$ denotes the modified Bessel function of the third kind with index 1 (see Abramowitz and Stegun 1970, Section 9.6), $\mu \in \mathbb{R}$ is the location parameter, $\delta > 0$ is the scale parameter, α and β, $0 \leq |\beta| \leq \alpha$, determines the shape of the density.

2.2.6 Reparametrized Johnson's S_U (JSU)

Rigby and Stasinopoulos (2005) reparametrize the S_U distribution of Johnson (1949) to make the mean equal to the parameter μ and the standard deviation equal to σ. Allows for skewness (through parameter ν – negative skewness when $\nu < 0$ and positive skewness when $\nu > 0$) and excess kurtosis (through the parameter $\tau > 0$). The density is given by:[8]

$$f_{JSU}(z_t|\mathscr{F}_{t-1}) = \frac{\tau}{c\sigma}\frac{1}{(r^2+1)^{1/2}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)$$

where

$$\begin{aligned} z \sim \mathcal{N}(0,1) &= -\nu + \tau \ln\left[r + (r^2+1)^{1/2}\right], \\ r &= \frac{z_t - (\mu + c\sigma w^{1/2}\sinh\Omega)}{c\sigma}, \\ c &= \left\{\frac{1}{2}(w-1)(w\cosh(2\Omega)+1\right\}^{-1/2}, \\ w &= \exp\left(\frac{1}{\tau^2}\right), \\ \Omega &= -\frac{\nu}{\tau}. \end{aligned}$$

3 Backtesting forecasts

In this section we will evaluate the volatility models by assessing one day out-of-sample forecasts of (i) volatility and (ii) value-at-risk (VaR), using the Swedish stock market price index OMX Stockholm 30 (OMXS30) as data. Daily closing prices, $\{P_t\}$, from 1990 to 2016[9] are used to calculate daily percentage log-returns, $r_t = 100\,(\ln P_t - \ln P_{t-1})$. The data is plotted in Figure 3.1, and Table 3.1 displays some summary statistics.

Table 3.1 Summary statistics for OMXS30 log returns. Q_k denotes the kth quartile, Q_2 is the median.

T	Mean	St. Dev.	Skew.	Exc. Kurt.	Min.	Q_1	Q_2	Q_3	Max.
6777	0.0294%	1.4707%	0.0987	4.0011	-8.8003%	-0.7369%	0.0574%	0.8003%	11.0228%

As seen in Table 3.1 the kurtosis is higher than for the normal distribution (excess kurtosis of 4 versus 0 for the normal), indicating that the true distribution is leptokurtic (meaning that is has heavier tails than a normal distribution), this can also be seen in Figure 3.2 which plots a histogram with normal and NIG[10] density curves overlayed, as well as quantile-quantile (Q-Q) plots, suggesting that the NIG distribution is a much better fit than a normal. We should note though,

[8] Taken from Stasinopoulos, Rigby, and Akantziliotou (2008).
[9] Obtained from `http://www.nasdaqomxnordic.com`.
[10] NIG is chosen to represent distributions that are able to capture excess kurtosis (and skewness), JSU and skew t could just as well have been used to illustrate the same point.

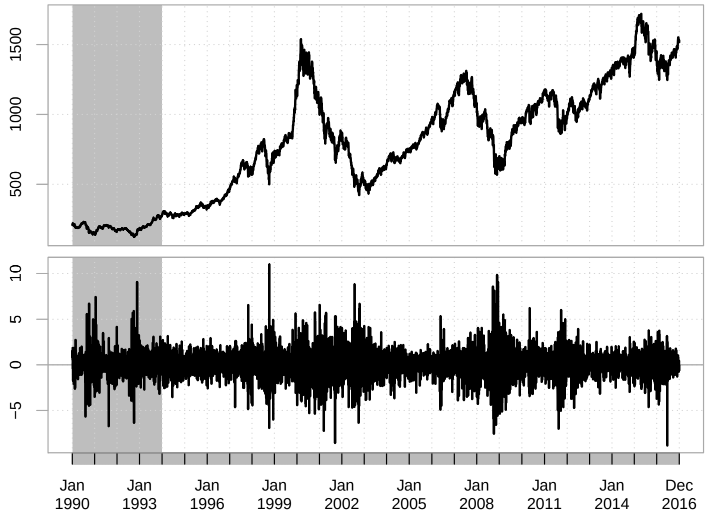

Figure 3.1 Daily closing prices and percentage log-returns of OMXS30 from 1990 to 2016. The training period is shaded and the out of sample period starts from 1993-12-28 and onwards.

that GARCH models allow for excess kurtosis even if the error terms are normally distributed; here follows the equation to calculate the kurtosis of the returns for a standard GARCH-N model (Taylor 2005, Eq. 9.13):

$$\text{kurtosis}(r_t) = \frac{3\,\mathrm{E}\left[\sigma_t^2\right]}{\sigma^4} = 3\left[\frac{1-(\alpha+\beta)^2}{1-2\alpha^2-(\alpha+\beta)^2}\right] > 3 \tag{3.1}$$

for $2\alpha^2 + (\alpha+\beta)^2 < 1$; otherwise, the kurtosis is infinite. For example, if $\alpha = 0.09$ and $\beta = 0.88$, then $\text{kurtosis}(r_t) = 4.13$, meaning that the excess kurtosis is 1.13. So GARCH-N can capture excess kurtosis, but the questions remains if it is flexible enough, especially for VaR forecasts when we are only interested in the left tail of the distribution.

3.1 Setup

The first 1000 days (the shaded area of Figure 3.1) is used as a training period leaving us with 5776 returns to be forecasted. The models are refitted every 5 days and—following Brownlees, Engle, and Kelly (2011a)—an expanding window is used (i.e., each estimation makes use of all available data up to that point, as opposed to a rolling window which uses a fixed number of observations for each estimation).[11] Plots of one day ahead forecasts of $\hat{\sigma}_t$ can be found in Appendix A,

[11] Brownlees, Engle, and Kelly (2011a) compares an expanding window with two different rolling windows, one short and one of medium length, and in their online appendix, Brownlees, Engle,

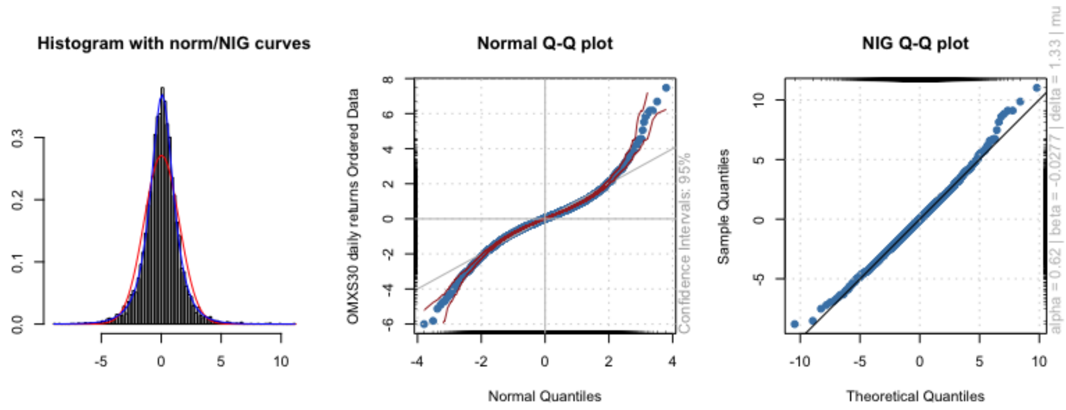

Figure 3.2 (Left) Histogram of the log returns with a normal curve (red) and NIG curve (blue). (Middle) Normal Q-Q plot with a 95% confidence band. (Right) NIG Q-Q plot.

in Figures A.1-A.6. Plots of the estimated parameters are found in Appendix B, Figures B.1-B.6.[12]

Before we start to discuss the results we should mention that the results for EGARCH-t, EGARCH-skew t and EGARCH-JSU are somewhat dubious. As can be seen in Figure B.5 the parameters for these models vary wildly compared to the other models, so the results from these models may probably not be relied upon, but we still leave them in and it is up to the reader to draw her own conclusions.

3.2 Backtesting volatility forecasts

The squared residual (squared demeaned return) $\hat{\varepsilon}_t^2 = (r_t - \hat{\mu})^2$ is used as a proxy of the true conditional variance σ_t^2, since the latter cannot be observed directly. Other alternatives include realized volatility (RV), but requires high frequency intra-day data. Both $\hat{\varepsilon}_t^2$ and RV are unbiased ex post proxies of conditional variance (Brownlees, Engle, and Kelly 2011a). See Brownlees, Engle, and Kelly (2011a,b) for a comparison of $\hat{\varepsilon}_t^2$ and RV (they actually use squared returns r_t^2 instead of $\hat{\varepsilon}_t^2$, but they do not use an intercept in their mean equation, assuming the mean to be zero, making $r_t = \varepsilon_t$, and in practice the daily returns are often very close to zero (see, e.g., Taylor 2005, Table 4.1)).

We use the two statistical loss functions used by Brownlees, Engle, and Kelly (2011a), namely the quasi-likelihood (QL) and the mean squared error (MSE).

and Kelly (2011b), they write: "Using a shorter, rolling estimation window tends to weaken forecasting accuracy. In some cases, the performance decreases by as much as 20%. The window length results are non-monotonic. While the full sample dominates, we often see that the medium estimation window does worse than the short window."

[12] The R (R Core Team 2017) package `rugarch` (Ghalanos 2015) is used to estimate the volatility models and `MCS` (Catania and Bernardi 2015) is used for the MCS procedure (see Section 3.2.1).

They argue that these two loss functions are the only proper loss functions to use among those commonly used to compare volatility forecast, since they belong to the class of Patton (2006) that "asymptotically generate the same ranking of models regardless of the proxy being used" and they also add that "This rank preservation holds as long as the proxy is unbiased and minimal regularity conditions are met. It ensures that model rankings achieved with proxies like squared residuals or realized volatility correspond to the ranking that would be achieved if forecasts were compared against the true volatility."

The loss functions are defined as follows:

$$\text{QL:} \quad L(\hat{\varepsilon}_t^2, \hat{\sigma}_t^2) = \ln\left(\hat{\sigma}_t^2\right) + \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}_t^2} \tag{3.2}$$

$$\text{MSE:} \quad L(\hat{\varepsilon}_t^2, \hat{\sigma}_t^2) = \left(\hat{\varepsilon}_t^2 - \hat{\sigma}_t^2\right)^2 \tag{3.3}$$

where $\hat{\sigma}_t^2$ is the volatility forecast from our model to be evaluated.

The means of the QL and MSE losses from the forecasts are displayed in Table 3.2. The table is ordered by the QL losses. The APARCH models are all ranked in the top by both QL and MSE. MSE then favors all the GJR-GARCH models, taking rank 7 to 12 and ranks the EGARCH models 13 to 18, while QL mixes EGARCH and GJR-GARCH models in the 7th to 18th spot. No EWMA, GARCH or IGARCH model makes the top half when ranking by either QL or MSE.

In the bottom we find the EWMA models when ranking by QL (spots 31-36, and 25-30 by MSE) and the IGARCH models are ranked the worst by MSE (spots 31-36, and 23-26, 29-30 by QL).

But how do we know that the best ranked model, APARCH-t, is statistically better than the worst one, EWMA-N (or IGARCH-N if ranked by MSE)? We will use a procedure developed by Hansen, Lunde, and Nason (2011) called the model confidence set (MCS).

3.2.1 The model confidence set (MCS)

The MCS procedure is a set of tests that eliminate models from the set $\mathcal{M}^0$ (all the considered models, in our case 36 models) one by one that do no pass a test of equal predictive ability (EPA) at a specified confidence level. A benchmark model to test against is not required when using the MCS procedure. The models that are not eliminated are included in the model confidence set (MCS) $\mathcal{M}^*$. See Hansen, Lunde, and Nason (2011) for the technical details and the `MCS` package (Catania and Bernardi 2015) for an R implementation.

The MCS procedure is used on our QL and MSE losses to compute two model confidence sets. Table 3.3 displays the results. For the MSE losses, all models

Table 3.2 Ranking of the volatility forecasts by QL and MSE (ordered by QL rank).

Model	QL		MSE	
	Value	Rank	Value	Rank
APARCH-t	1.46014	1	23.28331	1
APARCH-skew t	1.46030	2	23.28395	2
APARCH-JSU	1.46038	3	23.28732	3
APARCH-NIG	1.46048	4	23.29145	4
APARCH-N	1.46172	5	23.32723	5
APARCH-skew N	1.46186	6	23.33204	6
EGARCH-NIG	1.46336	7	23.44319	13
EGARCH-JSU	1.46343	8	23.45279	14
GJR-GARCH-t	1.46473	9	23.36620	8
EGARCH-N	1.46477	10	23.46315	15
GJR-GARCH-skew t	1.46483	11	23.36620	7
GJR-GARCH-JSU	1.46484	12	23.37059	9
EGARCH-skew N	1.46484	13	23.46372	16
GJR-GARCH-NIG	1.46488	14	23.37659	10
GJR-GARCH-N	1.46560	15	23.43196	11
GJR-GARCH-skew N	1.46569	16	23.43584	12
EGARCH-t	1.46737	17	23.66925	18
EGARCH-skew t	1.48524	18	23.54899	17
GARCH-JSU	1.48914	19	24.34193	21
GARCH-skew t	1.48915	20	24.34010	20
GARCH-t	1.48919	21	24.33996	19
GARCH-NIG	1.48920	22	24.34469	22
IGARCH-JSU	1.49001	23	24.59040	33
IGARCH-NIG	1.49003	24	24.60017	34
IGARCH-skew t	1.49011	25	24.58642	31
IGARCH-t	1.49013	26	24.58658	32
GARCH-skew N	1.49024	27	24.41989	24
GARCH-N	1.49045	28	24.41812	23
IGARCH-skew N	1.49129	29	24.72745	35
IGARCH-N	1.49144	30	24.73918	36
EWMA-t	1.50042	31	24.44000	25
EWMA-skew t	1.50043	32	24.44045	26
EWMA-JSU	1.50044	33	24.44076	27
EWMA-NIG	1.50047	34	24.44107	28
EWMA-skew N	1.50117	35	24.44800	30
EWMA-N	1.50125	36	24.44665	29

except for APARCH-t and APARCH-skew t are eliminated. Those are also the best ranked models when using the QL losses, with the difference that 12 other models also are included in the MCS—all the APARCH models are ranked in the top followed by a mix of EGARCH and GJR-GARCH models. We note that all EWMA, GARCH and IGARCH models are eliminated. Regarding the conditional distributions, we do not see any clear pattern, for MSE losses only models with two different distributions remain in the MCS: t and skew t; for QL: 3 models with JSU, 3 models with NIG, 2 models with skew t, 2 models with skew N, 2 models with t and 2 models with N distributions remain in the MCS.

Brownlees, Engle, and Kelly (2011a) conducts a similar study for forecasting one day ahead volatility using S&P500 daily returns from 1990 to 2008—with 2001-

Table 3.3 The model confidence sets for QL and MSE. All the EWMA, GARCH and IGARCH models were eliminated for both QL and MSE losses, and to save space they are not included in the table.

Model	Rank QL	Rank MSE
APARCH-t	1	1
APARCH-skew t	2	2
APARCH-JSU	3	Eliminated
APARCH-NIG	4	Eliminated
APARCH-N	5	Eliminated
APARCH-skew N	6	Eliminated
EGARCH-NIG	7	Eliminated
EGARCH-JSU	8	Eliminated
GJR-GARCH-t	9	Eliminated
EGARCH-N	10	Eliminated
GJR-GARCH-skew t	11	Eliminated
GJR-GARCH-JSU	12	Eliminated
GJR-GARCH-NIG	13	Eliminated
EGARCH-skew N	14	Eliminated
GJR-GARCH-N	Eliminated	Eliminated
GJR-GARCH-skew N	Eliminated	Eliminated
EGARCH-t	Eliminated	Eliminated
EGARCH-skew t	Eliminated	Eliminated

2009 as an out-of-sample period, comparing GARCH, GJR-GARCH, EGARCH, APARCH and NGARCH[13] with normal errors. They find[14] the QL losses smallest for GJR-GARCH, followed by APARCH, EGARCH, NGARCH and then GARCH; with the first four models performing better than GARCH on a 1% significance level using a Diebold-Mariano test (Diebold and Mariano 1995). They also repeat[15] the same procedure for nine U.S. sectoral equity indices and eighteen international equity indices.[16] For the equity sectors they find that APARCH performs the best, followed by GJR-GARCH, NGARCH, GARCH, and lastly EGARCH. APARCH, GJR-GARCH and NGARCH performs significantly better than GARCH on a 1% significance level. For the international equity indices they find that GJR-GARCH performs best, followed by NGARCH, APARCH, EGARCH and lastly GARCH, with the first three being significantly better than GARCH on a 1% significance level (but EGARCH not even on a 10% level). Their results agree with the current study in that GARCH performs poorly, but they have GJR-GARCH perform best for S&P500 and the international equity indices (but APARCH for the U.S. equity sectors), while GJR-GARCH-N was not even included in the model confidence set for the OMXS30 returns in the current study (Table 3.3) and where APARCH performed the best. They did, however, use a shorter out-of-sample period of

[13] Nonlinear GARCH (Engle 1990), not included in our study. With conditional variance $\sigma_t^2 = \omega + \alpha(\varepsilon_t + \gamma)^2 + \beta\sigma_{t-1}^2$.

[14] See Brownlees, Engle, and Kelly (2011b), Table 6.

[15] See Brownlees, Engle, and Kelly (2011b), Table 8.

[16] They also study ten exchange rates, which we, however, will not discuss here since it is not comparable to the data used in the current study.

2001-2008, compared to 1994-2016 in the current study, which could explain some differences.

Another study is Hansen and Lunde (2005) which compares 330 different GARCH specifications.[17] They study out-of-sample performance for IBM returns from June 1, 1999, to May 31, 2000, for a total of 254 trading days. They find APARCH$(2,2)$, specified with t-distributed errors and conditional mean of zero to perform the best. This is a comparable result to the current study where APARCH-t performed the best (even though we only include $(1,1)$ specifications for the conditional volatility and only a constant conditional mean).

3.3 Backtesting value-at-risk forecasts

3.3.1 Value-at-risk

The value-at-risk (VaR_t^α) is the loss that will exceed VaR with probability α, i.e., the conditional α quantile

$$\Pr\left(r_t < \text{VaR}_t^\alpha | \mathscr{F}_{t-1}\right) = \alpha \tag{3.4}$$

where $\mathscr{F}_{t-1}$ is the information set with all information available at time $t-1$.

We use the same data and models as when backtesting the volatility forecasts, and again we use the first 1000 days as a training period and the remaining 5776 days as an out-of-sample test period for one day ahead VaR-forecasts, where we calculate VaR_t^α for $\alpha = \{1\%, 5\%, 10\%\}$. The out-of-sample $\text{VaR}^{5\%}$ and $\text{VaR}^{1\%}$ one day forecasts for all models are plotted in Figures C.1-C.6 in Appendix C. Table 3.4 displays the actual and expected number of VaR exceedances for each of the models.

3.3.2 Christoffersen's conditional coverage test

In Table 3.5 VaR exceedances are used to calculate the conditional coverage of Christoffersen (1998), which tests the joint hypothesis that the frequency of exceedances is correct and that they are identically and independently distributed (iid). Christoffersen (1998) shows that the likelihood ratio test for conditional coverage is the sum:

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{3.5}$$

and LR_{cc} is distributed as $\chi^2_{(2)}$. LR_{uc} is a likelihood ratio test, and under the null hypothesis of correct unconditional coverage (frequency of exceedances) is

[17] 55 models for the conditional volatility, three different conditional mean specifications and either normal or t-distributed errors.

Table 3.4 The actual and expected number of VaR breaks, and actual divided by the total number of observations (5776).

	$\mathrm{VaR}^{1\%}$			$\mathrm{VaR}^{5\%}$			$\mathrm{VaR}^{10\%}$		
Model	actual	expected	$\frac{\text{actual}}{\text{obs.}}$	actual	expected	$\frac{\text{actual}}{\text{obs.}}$	actual	expected	$\frac{\text{actual}}{\text{obs.}}$
EWMA-N	105	57	1.82%	342	288	5.92%	584	577	10.11%
EWMA-t	77	57	1.33%	347	288	6.01%	622	577	10.77%
EWMA-skew N	95	57	1.64%	326	288	5.64%	573	577	9.92%
EWMA-skew t	73	57	1.26%	338	288	5.85%	617	577	10.68%
EWMA-NIG	68	57	1.18%	331	288	5.73%	615	577	10.65%
EWMA-JSU	68	57	1.18%	336	288	5.82%	618	577	10.70%
GARCH-N	75	57	1.30%	301	288	5.21%	538	577	9.31%
GARCH-t	63	57	1.09%	313	288	5.42%	598	577	10.35%
GARCH-skew N	73	57	1.26%	289	288	5.00%	526	577	9.11%
GARCH-skew t	61	57	1.06%	308	288	5.33%	592	577	10.25%
GARCH-NIG	60	57	1.04%	302	288	5.23%	589	577	10.20%
GARCH-JSU	60	57	1.04%	305	288	5.28%	592	577	10.25%
IGARCH-N	72	57	1.25%	282	288	4.88%	513	577	8.88%
IGARCH-t	58	57	1.00%	300	288	5.19%	575	577	9.95%
IGARCH-skew N	68	57	1.18%	272	288	4.71%	493	577	8.54%
IGARCH-skew t	52	57	0.90%	295	288	5.11%	570	577	9.87%
IGARCH-NIG	48	57	0.83%	282	288	4.88%	564	577	9.76%
IGARCH-JSU	48	57	0.83%	284	288	4.92%	568	577	9.83%
GJR-GARCH-N	72	57	1.25%	284	288	4.92%	543	577	9.40%
GJR-GARCH-t	55	57	0.95%	300	288	5.19%	578	577	10.01%
GJR-GARCH-skew N	73	57	1.26%	273	288	4.73%	534	577	9.25%
GJR-GARCH-skew t	55	57	0.95%	295	288	5.11%	573	577	9.92%
GJR-GARCH-NIG	54	57	0.93%	286	288	4.95%	570	577	9.87%
GJR-GARCH-JSU	55	57	0.95%	290	288	5.02%	572	577	9.90%
EGARCH-N	72	57	1.25%	293	288	5.07%	543	577	9.40%
EGARCH-t	57	57	0.99%	301	288	5.21%	586	577	10.15%
EGARCH-skew N	67	57	1.16%	286	288	4.95%	540	577	9.35%
EGARCH-skew t	30	57	0.52%	277	288	4.80%	585	577	10.13%
EGARCH-NIG	55	57	0.95%	297	288	5.14%	583	577	10.09%
EGARCH-JSU	56	57	0.97%	301	288	5.21%	586	577	10.15%
APARCH-N	74	57	1.28%	289	288	5.00%	547	577	9.47%
APARCH-t	56	57	0.97%	309	288	5.35%	581	577	10.06%
APARCH-skew N	72	57	1.25%	280	288	4.85%	544	577	9.42%
APARCH-skew t	54	57	0.93%	299	288	5.18%	582	577	10.08%
APARCH-NIG	53	57	0.92%	291	288	5.04%	580	577	10.04%
APARCH-JSU	54	57	0.93%	296	288	5.12%	581	577	10.06%

distributed as $\chi^2_{(1)}$,

$$LR_{uc} = -2\ln\left[(1-\alpha)^{T_0}\alpha^{T_1}\right] + 2\ln\left[(1-T_1/T)^{T_0}(T_1/T)^{T_1}\right] \quad (3.6)$$

where α is the known VaR$^{\alpha}$ coverage rate, T the total number of observations, T_0 the number of non-exceedances and T_1 the number of exceedances. LR_{ind} is the likelihood ratio test of independence, also distributed as $\chi^2_{(1)}$ under the null hypothesis,

$$LR_{ind} = -2\ln\left[(1-\hat{\pi}_2)^{(T_{00}+T_{11})}\hat{\pi}_2^{(T_{01}+T_{11})}\right] + 2\ln\left[(1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}\right] \quad (3.7)$$

where T_{ij} denotes the number of days where state j occurred after state i occurred the previous day, and i and j can take the value 1, which refers to exceedance or 0 which refers to non-exceedance; π_{ij} denotes the probability of state j given that the state the previous day was i, π_2 denotes the probability of exceedance given independence. Estimates of the probabilities are calculated as:

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}}, \qquad \hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}}, \qquad \hat{\pi}_2 = \frac{T_{01}+T_{11}}{T_{00}+T_{10}+T_{01}+T_{11}}$$

Table 3.5 displays the p-values for the Christoffersen likelihood ratio tests, both the joint test of conditional coverage (LR_{cc}), as well as the individual tests of unconditional coverage (LR_{uc}) and of independence (LR_{ind}).

For VaR$^{1\%}$, independence cannot rejected for any model. For UC and CC (the conclusions are the same for both) EWMA-N, EWMA-skew N and EGARCH-skew t are rejected at a 1% significance level, while in addition EWMA-t, GARCH-N and APARCH-N are rejected at a 5% significance level. At a 10% significance level, all the remaining models with normal distributions as well as GARCH-skew N, APARCH-skew N, GJR-GARCH-skew N and EWMA-skew t were also rejected. We note that no NIG or JSU distributed model can be rejected even at the 10% significance level, and only one t and two skew t distributed models are rejected, namely EWMA-t, EWMA-skew t and EGARCH-skew t (which probably had some convergence problems, see Figure B.5). Only two skew N distributed models cannot be rejected—EGARCH-skew N and IGARCH-skew N, while all N distributed models were rejected.

For VaR$^{5\%}$, all EWMA models are rejected at a 1% level for the joint CC test; no other model can be rejected—neither jointly nor by the the individual tests.

For VaR$^{10\%}$ and CC, IGARCH-skew N is rejected at a 1% level; in addition, EWMA-t, EWMA-skew t, EWMA-NIG, EWMA-JSU, GARCH-skew N and IGARCH-N are rejected at a 5% level.

In summary, for VaR$^{1\%}$, the distribution seems to matter more than the model specification, as a majority[18] of models that do not take excess kurtosis into

[18] On a 10% significance level.

account (those with N and skew N distributions) were rejected, while no NIG or JSU distributed model can be rejected. For $\text{VaR}^{5\%}$, however, EWMA performed very poorly and were all rejected, while none of the GARCH type models could be rejected. $\text{VaR}^{10\%}$ seems to be a blend, a majority of the EWMA models were rejected and the other three models to be rejected were N or skew N distributed.

Table 3.5 The conditional coverage test of Christoffersen (1998). LR_{uc} denotes the p-values from testing H_0: correct exceedances (equivalent to the test of Kupiec (1995)), LR_{ind} denotes the p-values from testing H_0: independent, and LR_{cc} denotes the p-values from testing the joint H_0: correct exceedances and independent.

	$\text{VaR}^{1\%}$			$\text{VaR}^{5\%}$			$\text{VaR}^{10\%}$		
Model	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
EWMA-N	0.000***	0.457	0.000***	0.002***	0.030**	0.001***	0.779	0.036**	0.106
EWMA-t	0.015**	0.149	0.019**	0.001***	0.008***	0.000***	0.054*	0.045**	0.021**
EWMA-skew N	0.000***	0.733	0.000***	0.028**	0.026**	0.008***	0.840	0.060*	0.167
EWMA-skew t	0.053*	0.172	0.060*	0.004***	0.012**	0.001***	0.087*	0.043**	0.030**
EWMA-NIG	0.188	0.203	0.187	0.013**	0.012**	0.002***	0.104	0.037**	0.030**
EWMA-JSU	0.188	0.203	0.187	0.005***	0.010**	0.001***	0.079*	0.046**	0.029**
GARCH-N	0.029**	0.160	0.035**	0.464	0.176	0.307	0.079*	0.292	0.123
GARCH-t	0.495	0.238	0.395	0.149	0.316	0.214	0.373	0.052*	0.102
GARCH-skew N	0.053*	0.172	0.060*	0.990	0.344	0.639	0.022**	0.341	0.046**
GARCH-skew t	0.671	0.254	0.476	0.251	0.251	0.268	0.529	0.064*	0.147
GARCH-NIG	0.768	0.262	0.510	0.429	0.186	0.305	0.618	0.094*	0.218
GARCH-JSU	0.768	0.262	0.510	0.332	0.217	0.291	0.529	0.064*	0.147
IGARCH-N	0.070*	0.178	0.078*	0.680	0.250	0.475	0.004***	0.477	0.012**
IGARCH-t	0.975	0.278	0.555	0.502	0.376	0.539	0.909	0.069*	0.189
IGARCH-skew N	0.188	0.203	0.187	0.306	0.531	0.487	0.000***	0.254	0.000***
IGARCH-skew t	0.438	0.331	0.462	0.709	0.303	0.549	0.738	0.067*	0.176
IGARCH-NIG	0.184	0.370	0.276	0.680	0.250	0.475	0.549	0.083*	0.186
IGARCH-JSU	0.184	0.370	0.276	0.771	0.275	0.529	0.673	0.058*	0.152
GJR-GARCH-N	0.070*	0.178	0.078*	0.771	0.774	0.920	0.126	0.993	0.310
GJR-GARCH-t	0.713	0.304	0.551	0.502	0.709	0.744	0.986	0.786	0.964
GJR-GARCH-skew N	0.053*	0.172	0.060*	0.336	0.550	0.526	0.053*	0.828	0.150
GJR-GARCH-skew t	0.713	0.304	0.551	0.709	0.607	0.817	0.840	0.900	0.972
GJR-GARCH-NIG	0.615	0.313	0.529	0.866	0.441	0.732	0.738	0.969	0.945
GJR-GARCH-JSU	0.713	0.304	0.551	0.942	0.511	0.804	0.806	0.923	0.966
EGARCH-N	0.070*	0.178	0.078*	0.800	0.277	0.536	0.126	0.652	0.280
EGARCH-t	0.920	0.286	0.564	0.464	0.730	0.721	0.713	0.947	0.933
EGARCH-skew N	0.233	0.210	0.224	0.866	0.614	0.868	0.096*	0.590	0.216
EGARCH-skew t	0.000***	0.576	0.000***	0.473	0.934	0.771	0.746	0.855	0.933
EGARCH-NIG	0.713	0.304	0.551	0.622	0.473	0.685	0.813	0.869	0.959
EGARCH-JSU	0.815	0.295	0.562	0.464	0.392	0.530	0.713	0.825	0.912
APARCH-N	0.040**	0.166	0.046**	0.990	0.493	0.791	0.176	0.856	0.394
APARCH-t	0.815	0.295	0.562	0.228	0.382	0.329	0.882	0.937	0.986
APARCH-skew N	0.070*	0.178	0.078*	0.593	0.346	0.556	0.137	0.788	0.319
APARCH-skew t	0.615	0.313	0.529	0.540	0.361	0.546	0.847	0.960	0.980
APARCH-NIG	0.523	0.322	0.499	0.894	0.374	0.668	0.916	0.913	0.989
APARCH-JSU	0.615	0.313	0.529	0.665	0.317	0.552	0.882	0.937	0.986

Note: $p < 0.10$ is denoted by *, $p < 0.05$ by **, and $p < 0.01$ by ***.

3.3.3 Statistical losses

One drawdown of the Christoffersen (1998) test is that it does not take the size of each VaR break into account. Therefore, we will calculate the statistical losses for VaR_t^α, using the asymmetric VaR loss function of González-Rivera, Lee, and Mishra (2004):

$$L(r_t, \text{VaR}_t^\alpha) = (\alpha - \mathbb{1}_{\{r_t < \text{VaR}_t^\alpha\}})(r_t - \text{VaR}_t^\alpha), \tag{3.8}$$

where $\mathbb{1}_{\{r_t < \text{VaR}_t^\alpha\}}$ is the indicator function for VaR exceedances, which takes the value 1 when $r_t < \text{VaR}_t^\alpha$ and 0 otherwise. It penalizes the losses below the α quantile level, i.e., $r_t < \text{VaR}_t^\alpha$, with weight $1 - \alpha$. Table 3.6 ranks the models by the sum of the statistical losses.

Table 3.6 Ranking of the models by the sum of the statistical losses for $\text{VaR}^{1\%}$, $\text{VaR}^{5\%}$ and $\text{VaR}^{10\%}$.

Rank	$\text{VaR}^{1\%}$	$\text{VaR}^{5\%}$	$\text{VaR}^{10\%}$
1	EGARCH-NIG	APARCH-JSU	GJR-GARCH-JSU
2	EGARCH-JSU	APARCH-skew t	GJR-GARCH-NIG
3	EGARCH-skew N	APARCH-NIG	GJR-GARCH-skew t
4	APARCH-skew t	APARCH-t	GJR-GARCH-t
5	APARCH-JSU	GJR-GARCH-skew t	APARCH-skew t
6	APARCH-NIG	GJR-GARCH-JSU	APARCH-JSU
7	GJR-GARCH-skew t	GJR-GARCH-NIG	APARCH-NIG
8	GJR-GARCH-JSU	EGARCH-NIG	APARCH-t
9	GJR-GARCH-NIG	GJR-GARCH-t	GJR-GARCH-N
10	APARCH-t	APARCH-skew N	GJR-GARCH-skew N
11	APARCH-skew N	GJR-GARCH-skew N	APARCH-skew N
12	EGARCH-N	APARCH-N	APARCH-N
13	GJR-GARCH-t	GJR-GARCH-N	EGARCH-NIG
14	GJR-GARCH-skew N	EGARCH-JSU	EGARCH-JSU
15	APARCH-N	EGARCH-skew N	EGARCH-skew N
16	GJR-GARCH-N	EGARCH-N	EGARCH-N
17	EGARCH-t	EGARCH-t	EGARCH-t
18	IGARCH-skew N	EGARCH-skew t	EGARCH-skew t
19	GARCH-skew N	GARCH-NIG	GARCH-NIG
20	IGARCH-N	GARCH-JSU	GARCH-skew t
21	GARCH-NIG	GARCH-skew t	GARCH-JSU
22	GARCH-JSU	IGARCH-skew t	IGARCH-skew t
23	GARCH-skew t	IGARCH-t	IGARCH-JSU
24	GARCH-t	IGARCH-JSU	IGARCH-t
25	GARCH-N	IGARCH-NIG	GARCH-t
26	IGARCH-t	GARCH-skew N	IGARCH-NIG
27	IGARCH-skew t	GARCH-t	GARCH-N
28	IGARCH-JSU	GARCH-N	GARCH-skew N
29	IGARCH-NIG	IGARCH-N	EWMA-N
30	EGARCH-skew t	IGARCH-skew N	EWMA-skew N
31	EWMA-JSU	EWMA-skew N	EWMA-NIG
32	EWMA-NIG	EWMA-NIG	EWMA-JSU
33	EWMA-skew t	EWMA-JSU	EWMA-skew t
34	EWMA-t	EWMA-skew t	EWMA-t
35	EWMA-skew N	EWMA-N	IGARCH-N
36	EWMA-N	EWMA-t	IGARCH-skew N

The MCS procedure of Hansen, Lunde, and Nason (2011) described in Section 3.2.1 is used again to produce the model confidence sets for $\text{VaR}^{1\%}$, $\text{VaR}^{5\%}$ and $\text{VaR}^{10\%}$ losses. As seen Table 3.7, for all levels of α, all the EWMA, IGARCH and GARCH models are again eliminated just as in the MCS for the volatility forecasts. For some reason, EGARCH-t and EGARCH-skew t are again also eliminated for all three levels of α.[19]

When looking at the ranking of the non-eliminated models (i.e., those in the MCS) we see some differences for different values of α.

For $\text{VaR}^{1\%}$, 16 models are included in the MCS. The EGARCH and APARCH models that have distributions with skewness (skew-N, skew-t, NIG and JSU) seems to be preferred and are in the top six. Then comes the three GJR-GARCH models that allows for both skewness and excess kurtosis and then a mixture of the rest.

For $\text{VaR}^{5\%}$, 15 models are included in the MCS, in addition to the above mentioned models eliminated for all levels of VaR, EGARCH-N is also eliminated. APARCH that have distributions that allows for skewness and excess kurtosis (JSU, skew-t and NIG) are in the top three, then we have APARCH-t at four, followed by the three GJR models allowing for skewness and excess kurtosis (skew-t, JSU, NIG). After that we have a mixture of the remaining models.

For $\text{VaR}^{10\%}$, 10 models are included in the MCS—in addition to EWMA, IGARCH and GARCH, all the EGARCH models, APARCH-N and APARCH-skew N are also eliminated. All the GJR-GARCH models remain and the ones that allows for skewness and excess kurtosis forms the top three (JSU, NIG, skew-t) and then comes GJR-GARCH-t, after which we have the four remaining APARCH models, and finishing off with GJR-GARCH-skew N and N.

Table 3.7 The model confidence sets for $\text{VaR}^{1\%}$, $\text{VaR}^{5\%}$ and $\text{VaR}^{10\%}$.

Rank	$\text{VaR}^{1\%}$	$\text{VaR}^{5\%}$	$\text{VaR}^{10\%}$
1	EGARCH-NIG	APARCH-JSU	GJR-GARCH-JSU
2	EGARCH-JSU	APARCH-skew t	GJR-GARCH-NIG
3	APARCH-skew t	APARCH-NIG	GJR-GARCH-skew t
4	EGARCH-skew N	APARCH-t	GJR-GARCH-t
5	APARCH-JSU	GJR-GARCH-skew t	APARCH-skew t
6	APARCH-NIG	GJR-GARCH-JSU	APARCH-JSU
7	GJR-GARCH-skew t	GJR-GARCH-NIG	APARCH-NIG
8	GJR-GARCH-JSU	EGARCH-NIG	APARCH-t
9	GJR-GARCH-NIG	GJR-GARCH-t	GJR-GARCH-skew N
10	EGARCH-N	APARCH-skew N	GJR-GARCH-N
11	APARCH-t	EGARCH-JSU	-
12	APARCH-skew N	GJR-GARCH-skew N	-
13	GJR-GARCH-skew N	GJR-GARCH-N	-
14	GJR-GARCH-t	APARCH-N	-
15	APARCH-N	EGARCH-skew N	-
16	GJR-GARCH-N	-	-

[19] We suspect convergence issues with the solver, which also seems to be an issue with EGARCH-JSU, see the parameter instability in Figure B.5.

González-Rivera, Lee, and Mishra (2004) analyzes $\text{VaR}^{5\%}$ forecasts using conditional volatility models with normal errors for S&P500 returns from April 1, 1970 to November 17, 2000, with the 999 last returns used as an out-of-sample period. And similar to the current study, they find that EWMA and IGARCH performs very poorly, while they find the stochastic volatility model[20] (not included in our study) to perform the best. Amongst the GARCH models included, EGARCH performs the best, which does not agree with our findings, since EGARCH-N was not included in the MCS for $\text{VaR}^{5\%}$.[21]

4 Conclusions

For forecasting one day ahead volatility the model specification seems to be more important than the conditional distribution, and the winner is APARCH with the only two models left in the MCS for the MSE losses and with all the top six models in the MCS with the QL losses, while EGARCH and GJR-GARCH make up the remaining models. No EWMA, GARCH or IGARCH models remain at all.

For forecasting one day ahead value-at-risk the conditional distributions are more important than for forecasting volatility, but the model specification is still important, with all the EWMA, GARCH and IGARCH models yet again eliminated from the models confidence sets for all levels of VaR. Models distributed as JSU, NIG and skew t (i.e., the distributions that allow for both excess kurtosis and skewness) top all three levels of VaR. However, different model specifications seems to be better for different levels of VaR, with EGARCH in the top for $\text{VaR}^{1\%}$, APARCH in top for $\text{VaR}^{5\%}$ and GJR-GARCH in top for $\text{VaR}^{10\%}$.

4.1 Suggestions for further research

There are many ways that this study could be extended. We could:

- Extend the scope by (i) including even more models and conditional distributions, (ii) by including stochastic volatility models (Taylor 1982, 1986) or implied volatilities from option prices.
- Include longer time horizons for the forecast, instead of only focusing on one day ahead, such as forecasting a week and a month ahead.

[20] See e.g., Shephard (1996, 2005) and Taylor (1994).

[21] Only two models with normal errors were included in the MCS for $\text{VaR}^{5\%}$ in our study, namely GJR-GARCH-N and APARCH-N and they were ranked in the bottom, so the results of González-Rivera, Lee, and Mishra (2004) may not be completely comparable to our study since they only use normal errors to model the conditional volatilities.

- Use more data. The current study only uses one equity index return series. We could expand the study by including data from multiple asset classes and countries.
- Use other proxies of volatility—since squared returns are quite noisy—e.g., realized volatilities or daily range.

References

Abramowitz, Milton and Irene Stegun, eds. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. U.S. Government Printing Office, Washington, D.C.: National Bureau of Standards Applied Mathematics Series.

Andersen, T. G., R. A. Davis, et al., eds. (2009). *Handbook of Financial Time Series*. Berlin and Heidelberg: Springer-Verlag. ISBN: 978-3-540-71296-1.

Andersen, Torben G., Tim Bollerslev, et al. (2006). "Volatility Forecasting." In: *Handbook of Economic Forecasting*. Ed. by Graham Elliott, Clive W.J. Granger, and Allan Timmermann. Vol. 1. Elsevier. Chap. 15. DOI: 10.1016/S1574-0706(05)01015-3.

Barndorff-Nielsen, Ole E. (1997). "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling." In: *Scandinavian Journal of Statistics* 24.1, pp. 1–13.

Black, Fischer (1976). "The pricing of commodity contracts." In: *Journal of Financial Economics* 3.1–2, pp. 167–179.

Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroskedasticity." In: *Journal of Econometrics* 31.3, pp. 307–327.

— (1987). "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." In: *The Review of Economics and Statistics* 69.3, pp. 542–547.

— (2008). *Glossary to ARCH (GARCH)*. CREATES Research Papers 2008-49. School of Economics and Management, University of Aarhus. DOI: 10.2139/ssrn.1263250.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner (1992). "ARCH modeling in finance: A review of the theory and empirical evidence." In: *Journal of Econometrics* 52, pp. 5–59.

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson (1994). "ARCH Models." In: *Handbook of Econometrics*. Ed. by Robert F. Engle and D.L. McFadden. Vol. IV. Elsevier Science. Chap. 49, pp. 2959–3038.

Brownlees, Christian, Robert Engle, and Bryan Kelly (2011a). "A Practical Guide to Volatility Forecasting Through Calm and Storm." In: *Journal of Risk* 14.2, pp. 1–20.

— (2011b). *A Practical Guide to Volatility Forecasting Through Calm and Storm (Appendix)*. Tech. rep. URL: http://84.89.132.1/~cbrownlees/docs/volfor-appendix.pdf.

Catania, Leopoldo and Mauro Bernardi (2015). *MCS: Model Confidence Set Procedure*. R package version 0.1.1.

Christoffersen, Peter (1998). "Evaluating Interval Forecasts." In: *International Economic Review* 39.4, pp. 841–862.

Diebold, Francis X and Roberto S Mariano (1995). "Comparing Predictive Accuracy." In: *Journal of Business & Economic Statistics* 13.3, pp. 253–63.

Ding, Z., Clive W. J. Granger, and Robert F. Engle (1993). "A Long Memory property of stock market returns and a new model." In: *Journal of Empirical Finance* 1, pp. 83–106.

Dowd, Kevin (2005). *Measuring market risk.* Second edition. Wiley Finance. John Wiley & Sons.

Engle, Robert F. (1982). "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of UK Inflation." In: *Econometrica* 50.4, pp. 987–1008. JSTOR: 1912773.

— (1990). "Discussion: stock market volatility and the crash of '87." In: *Review of Financial Studies* 3, pp. 103–106.

Engle, Robert F. and Tim Bollerslev (1986). "Modelling the persistence of conditional variances." In: *Econometric Reviews,* 5, pp. 1–50.

Fernandez, Carmen and Mark F. J. Steel (1998). "On Bayesian Modeling of Fat Tails and Skewness." In: *Journal of the American Statistical Association* 93.441, pp. 359–371. ISSN: 01621459.

Ghalanos, Alexios (2015). *rugarch: Univariate GARCH models.* R package version 1.3-6.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." In: *Journal of Finance* 48.5, pp. 1779–1801.

González-Rivera, Gloria, Tae-Hwy Lee, and Santosh Mishra (2004). "Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood." In: *International Journal of forecasting* 20.4, pp. 629–645.

Granger, Clive W. J. and Zhuanxin Ding (1995). "Some Properties of Absolute Return: An Alternative Measure of Risk." In: *Annales d'Économie et de Statistique* 40, pp. 67–91. DOI: 10.2307/20076016.

Hansen, Peter Reinhard and Asger Lunde (2005). "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?" In: *Journal of Applied Econometrics* 20.7, pp. 873–889.

Hansen, Peter Reinhard, Asger Lunde, and James M Nason (2011). "The model confidence set." In: *Econometrica* 79.2, pp. 453–497.

Hentschel, Ludger (1995). "All in the family nesting symmetric and asymmetric GARCH models." In: *Journal of Financial Economics* 39.1, pp. 71–104.

Johnson, N. L. (1949). "Systems of Frequency Curves Generated by Methods of Translation." In: *Biometrika* 36.1/2, pp. 149–176. ISSN: 00063444. DOI: 10.2307/2332539.

J.P. Morgan and Reuters (1996). *RiskMetrics: Technical Document (fourth edition)*. Tech. rep. New York: Morgan Guaranty Trust Company. URL: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Kupiec, Paul H. (1995). "Techniques for verifying the accuracy of risk measurement models." In: *The journal of Derivatives* 3.2, pp. 73–84.

Nelson, Daniel B. (1991). "Conditional heteroskedasticity in asset returns: A new approach." In: *Econometrica: Journal of the Econometric Society*, pp. 347–370.

Patton, A. J. and K. Sheppard (2009). "Evaluating Volatility and Correlation Forecasts." In: *Hanbook of Financial Time Series*. Ed. by T. G. Andersen et al. Berlin and Heidelberg: Elsevier, pp. 801–838. ISBN: 978-3-540-71296-1.

Patton, Andrew J. (2006). *Volatility Forecast Comparison using Imperfect Volatility Proxies*. Research Paper 175. Sydney: Quantitative Finance Research Centre, University of Technology.

— (2011). "Volatility forecast comparison using imperfect volatility proxies." In: *Journal of Econometrics* 160.1, pp. 246–256. ISSN: 0304-4076. DOI: http://dx.doi.org/10.1016/j.jeconom.2010.03.034.

Poon, Ser-Huang and Clive W.J. Granger (2003). "Forecasting Volatility in Financial Markets: A Review." In: *Journal of Economic Literature* 41.2, pp. 478–539.

— (2005). "Practical issues in forecasting volatility." In: *Financial Analysts Journal* 61.1, pp. 45–56.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.r-project.org.

Rigby, R. A. and D. M. Stasinopoulos (2005). "Generalized additive models for location, scale and shape." In: *Journal of the Royal Statistical Society Series C* 54.3, pp. 507–554.

Shephard, Neil (1996). "Statistical aspects of ARCH and Stochastic Volatility." In: *Time Series Models in Econometrics, Finance and Other Fields*. Ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndoff-Nielsen. London: Chapman & Hall. Chap. 1, pp. 1–67.

— ed. (2005). *Stochastic Volatility. Selected Readings*. Oxford and New York: Oxford University Press. ISBN: 0199257205.

Stasinopoulos, Mikis, Bob Rigby, and Calliope Akantziliotou (2008). *Instructions on how to use the gamlss package in R (Second Edition)*. Tech. rep. URL: http://www.gamlss.org/wp-content/uploads/2013/01/gamlss-manual.pdf.

Taylor, Stephen J. (1982). "Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices, 1961–79. Theory and Practice 1." In: *Time Series Analysis*. Ed. by O. D. Anderson. Amsterdam: North-Holland, pp. 203–226.

— (1986). *Modelling Financial Time Series*. Chichester: Wiley.

Taylor, Stephen J. (1994). "Modeling Stochastic Volatility: A Review and Comparative Study." In: *Mathematical Finance* 4.2, pp. 183–204.

— (2005). *Asset Price Dynamics, Volatility and Prediction.* Princeton and Oxford: Princeton University Press. ISBN: 0-691-11537-0.

Teräsvirta, Timo (2009). "An introduction to univariate GARCH models." In: *Handbook of Financial Time Series.* Ed. by T. G. Andersen et al. Berlin and Heidelberg: Springer-Verlag, pp. 17–42. ISBN: 978-3-540-71296-1.

Zakoian, Jean-Michel (1994). "Threshold heteroskedastic models." In: *Journal of Economic Dynamics and control* 18.5, pp. 931–955.

Appendix A Plots of one day ahead forecasts of $\hat{\sigma}_t$

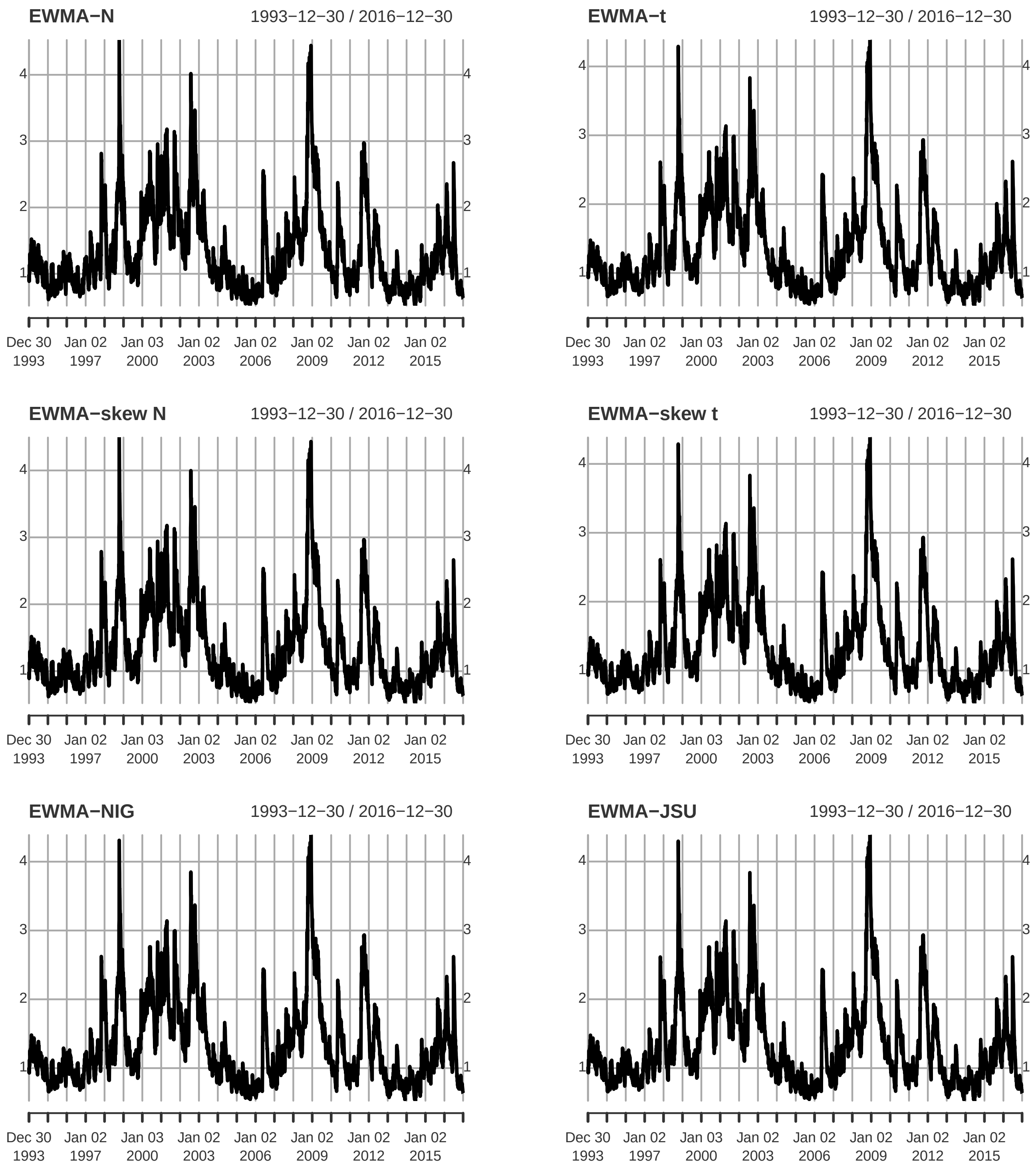

Figure A.1 One day ahead forecasts of $\hat{\sigma}_t$ for EWMA.

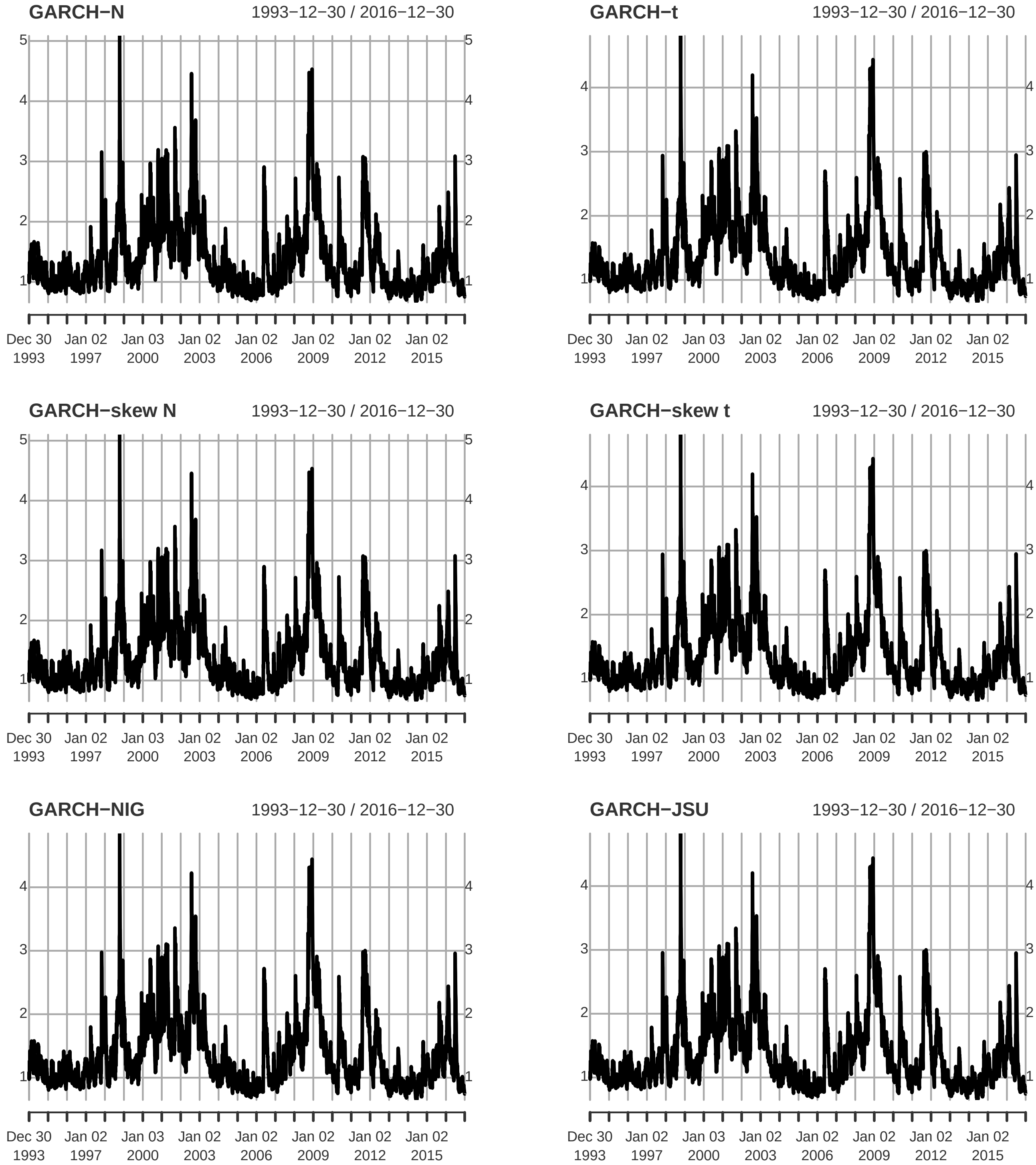

Figure A.2 One day ahead forecasts of $\hat{\sigma}_t$ for GARCH.

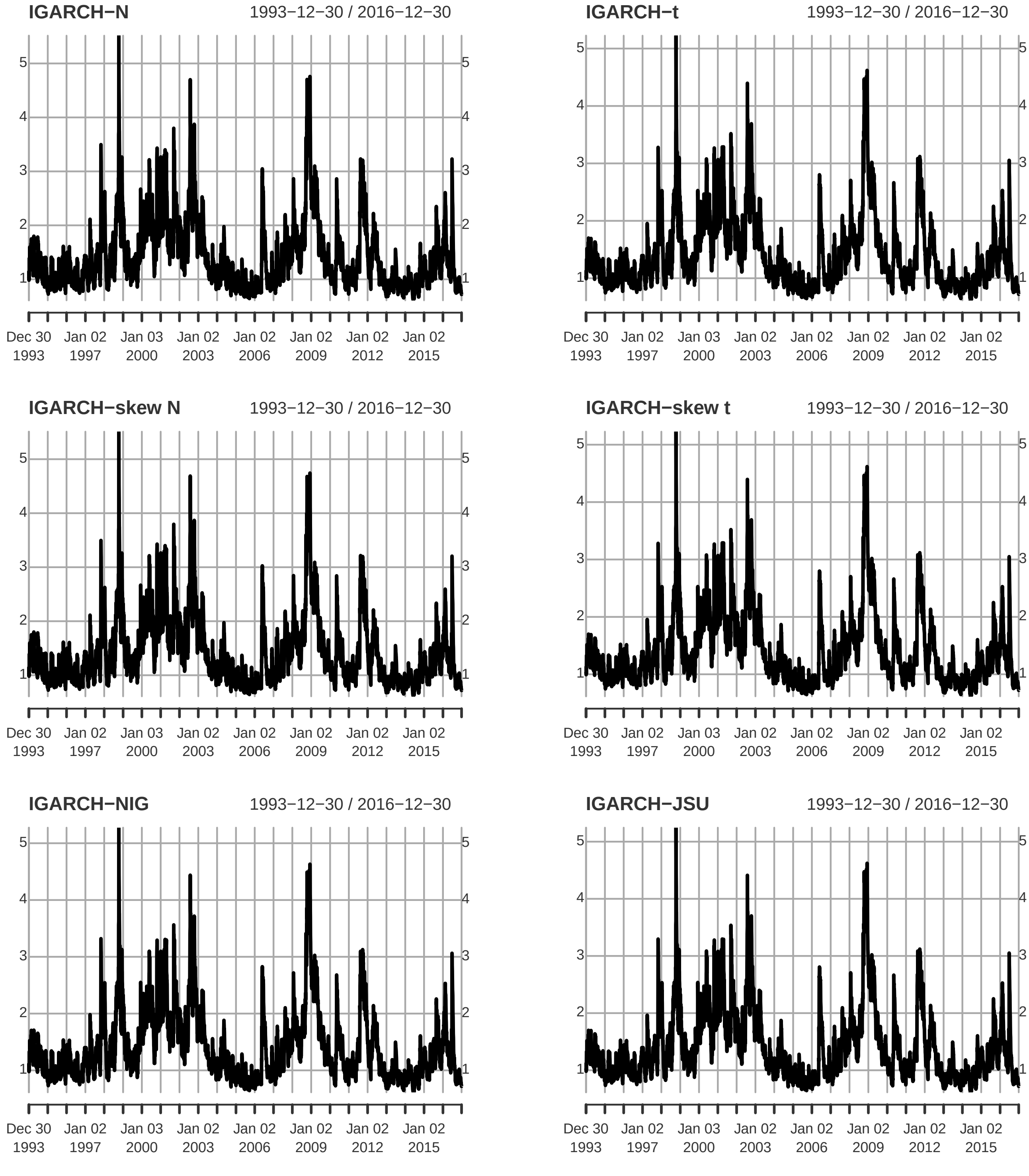

Figure A.3 One day ahead forecasts of $\hat{\sigma}_t$ for IGARCH.

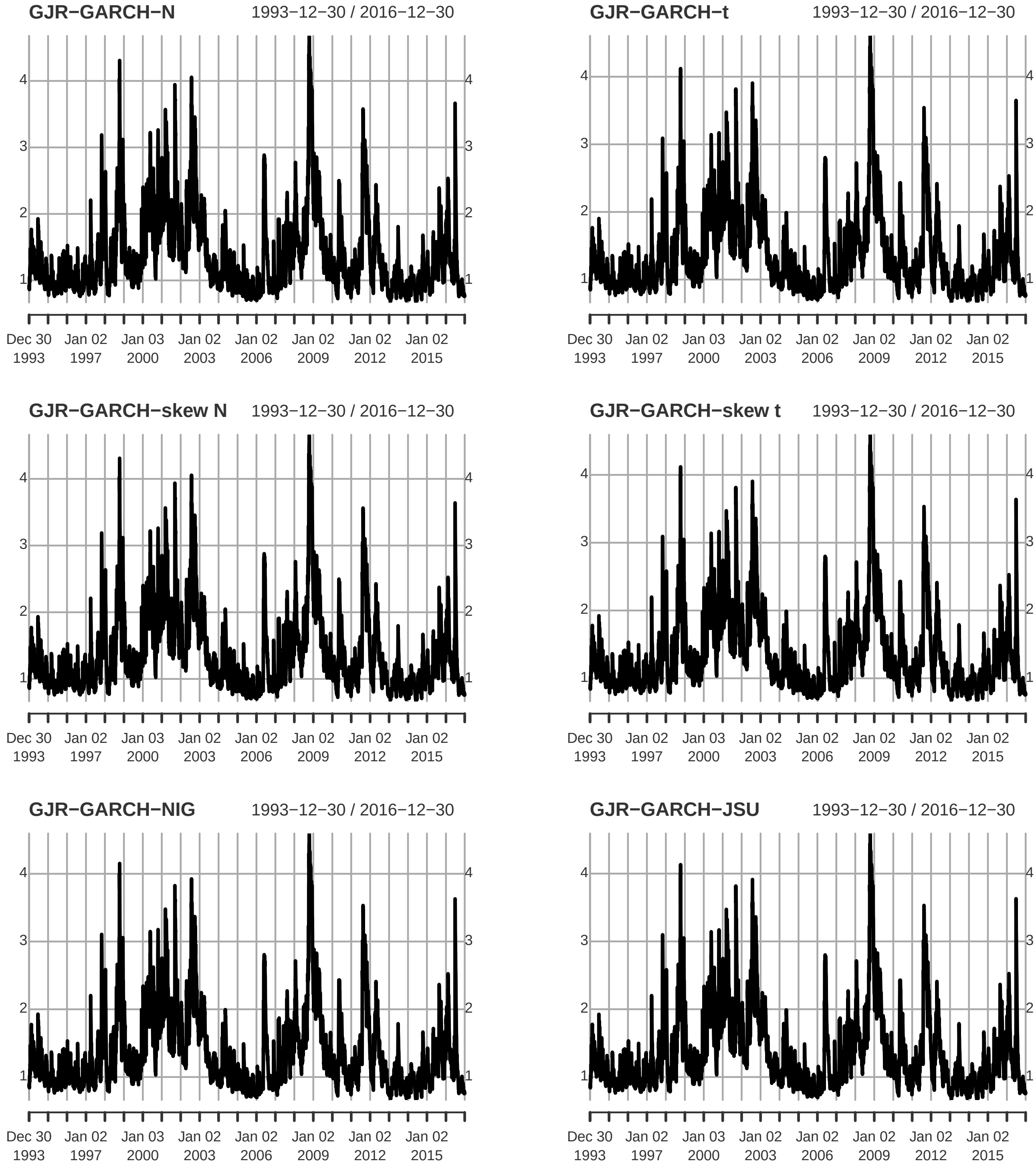

Figure A.4 One day ahead forecasts of $\hat{\sigma}_t$ for GJR-GARCH.

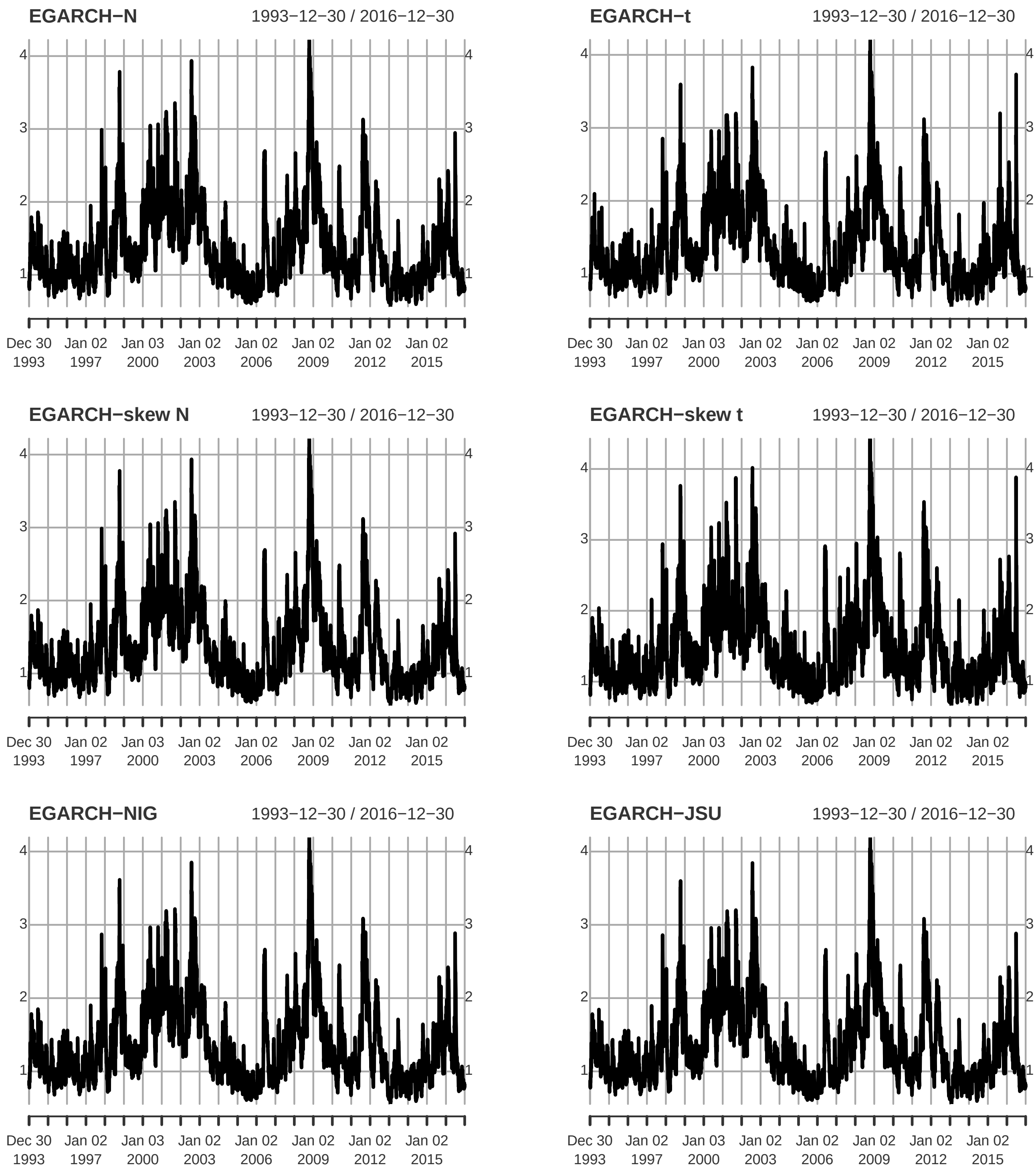

Figure A.5 One day ahead forecasts of $\hat{\sigma}_t$ for EGARCH.

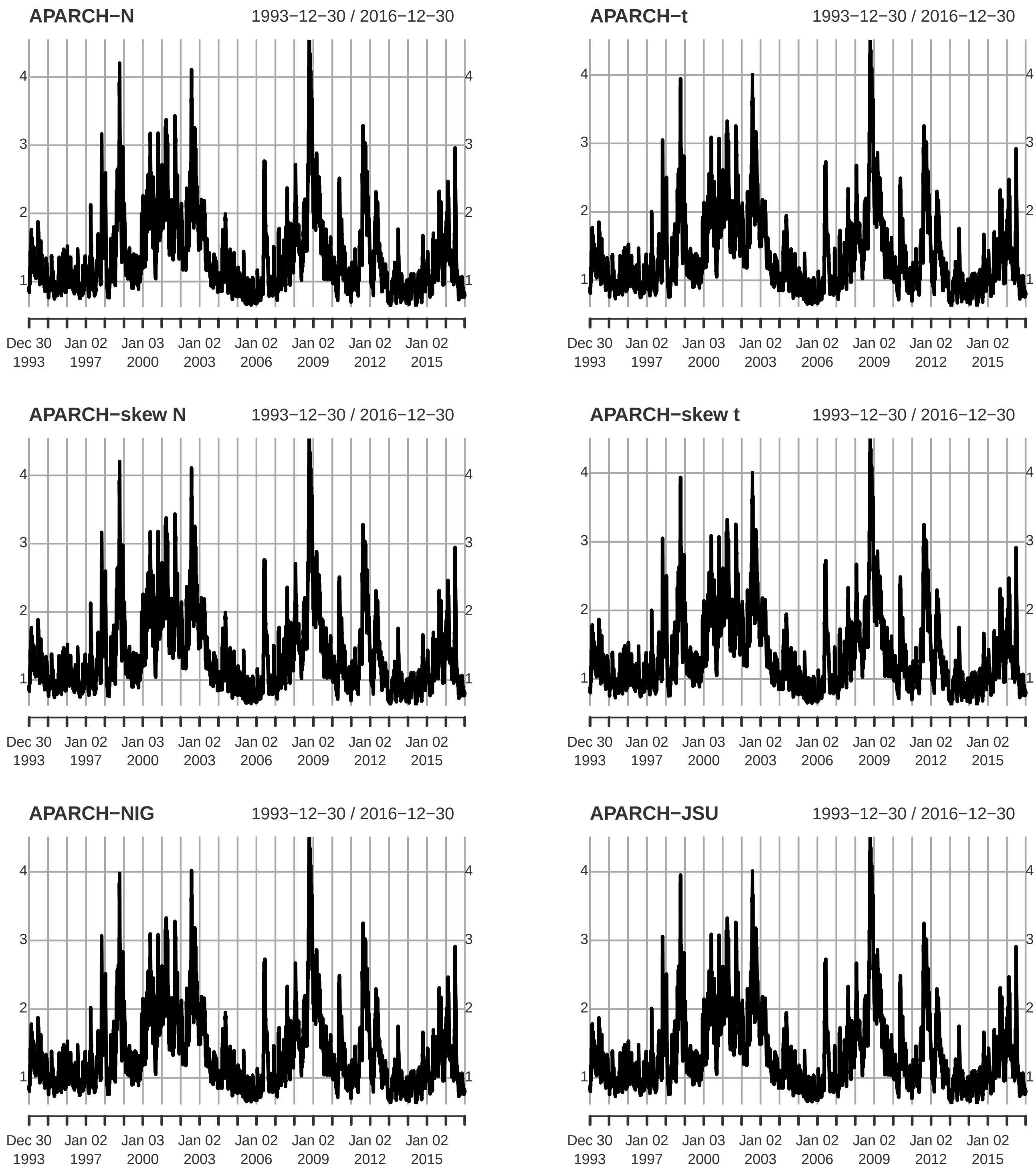

Figure A.6 One day ahead forecasts of $\hat{\sigma}_t$ for APARCH.

Appendix B Parameter plots

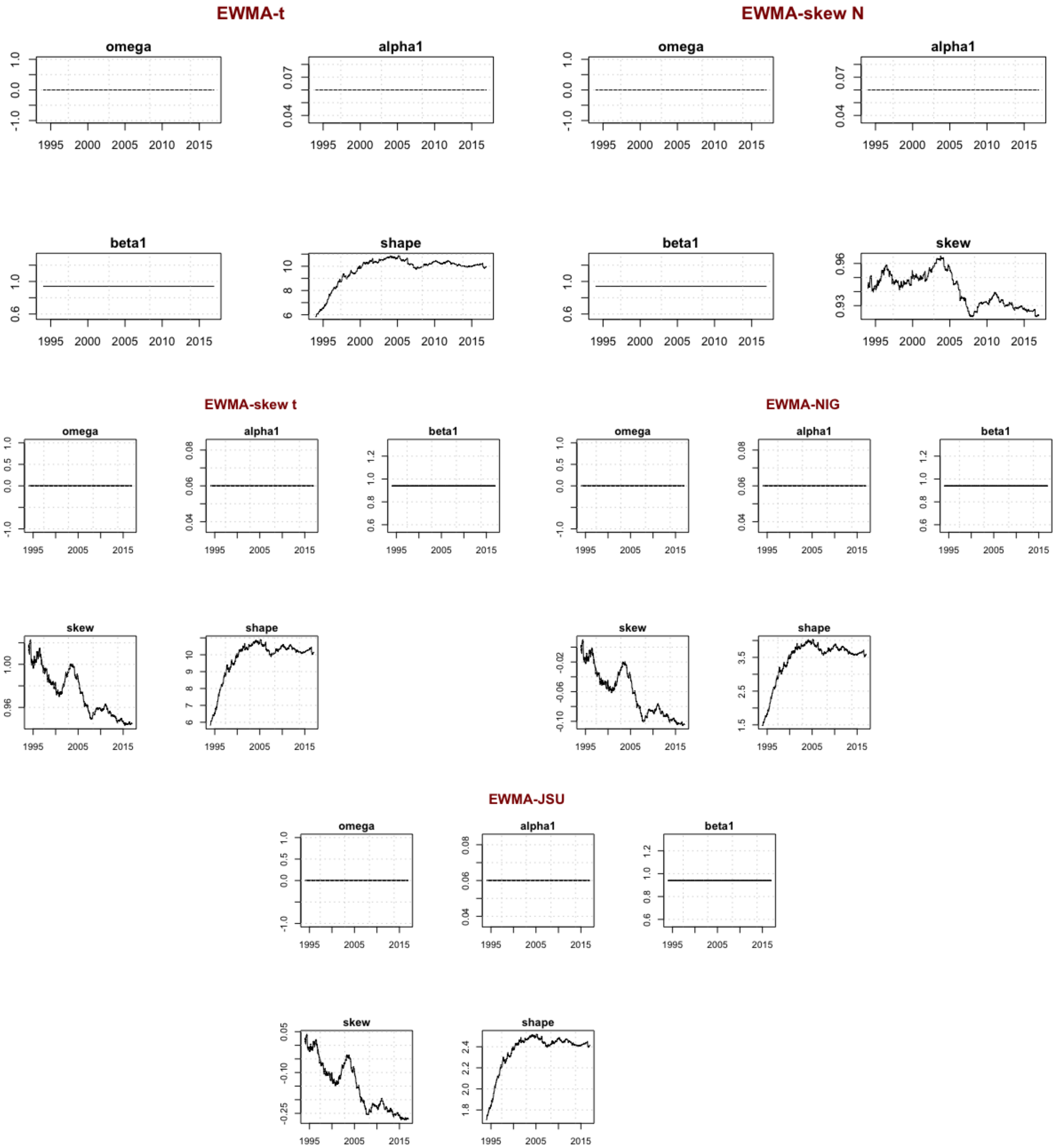

Figure B.1 Parameter plots for all the estimations of the EWMA models. For EWMA-N all parameters are constants so they are not included.

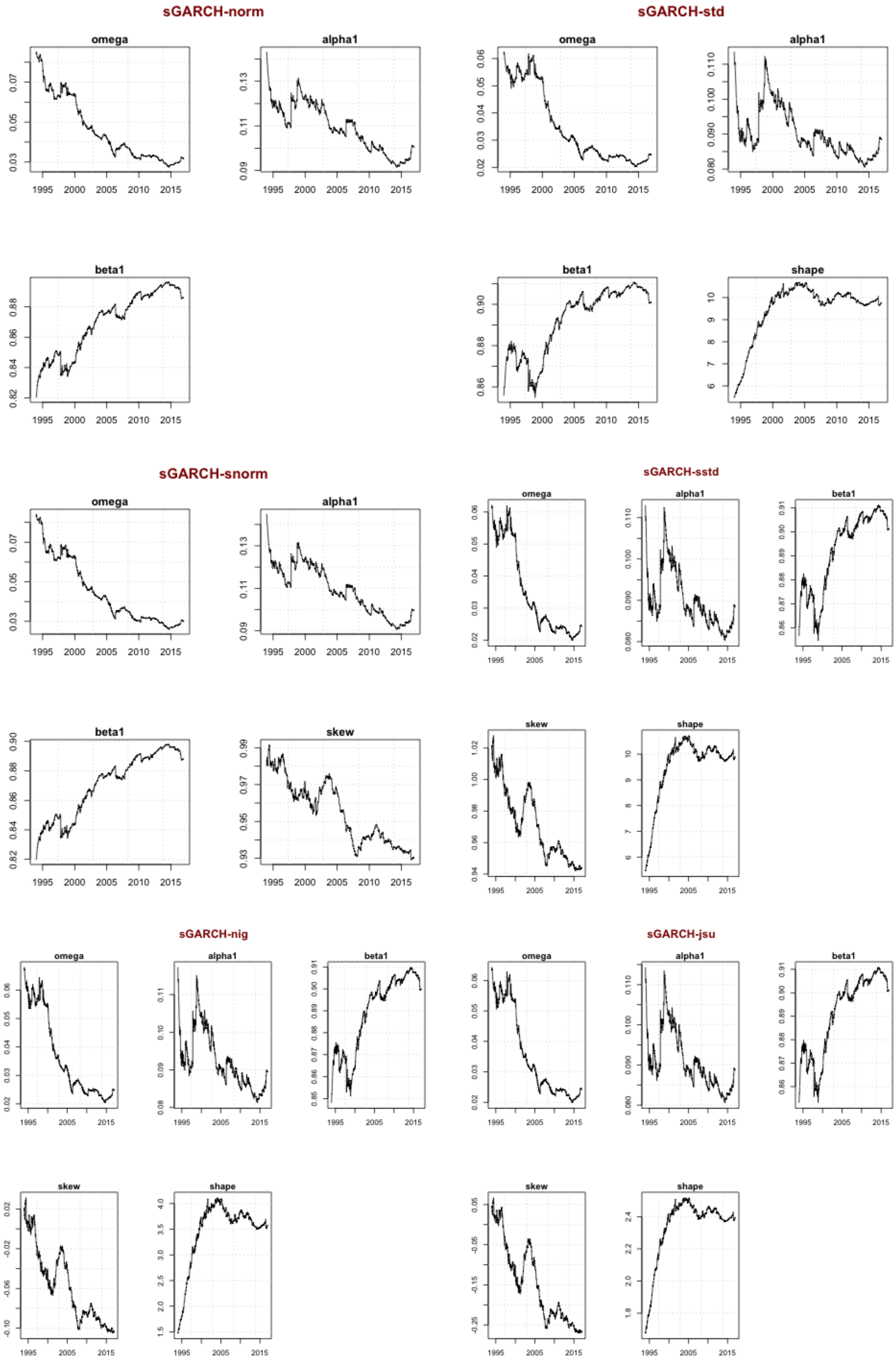

Figure B.2 Parameter plots for all the estimations of the GARCH models.

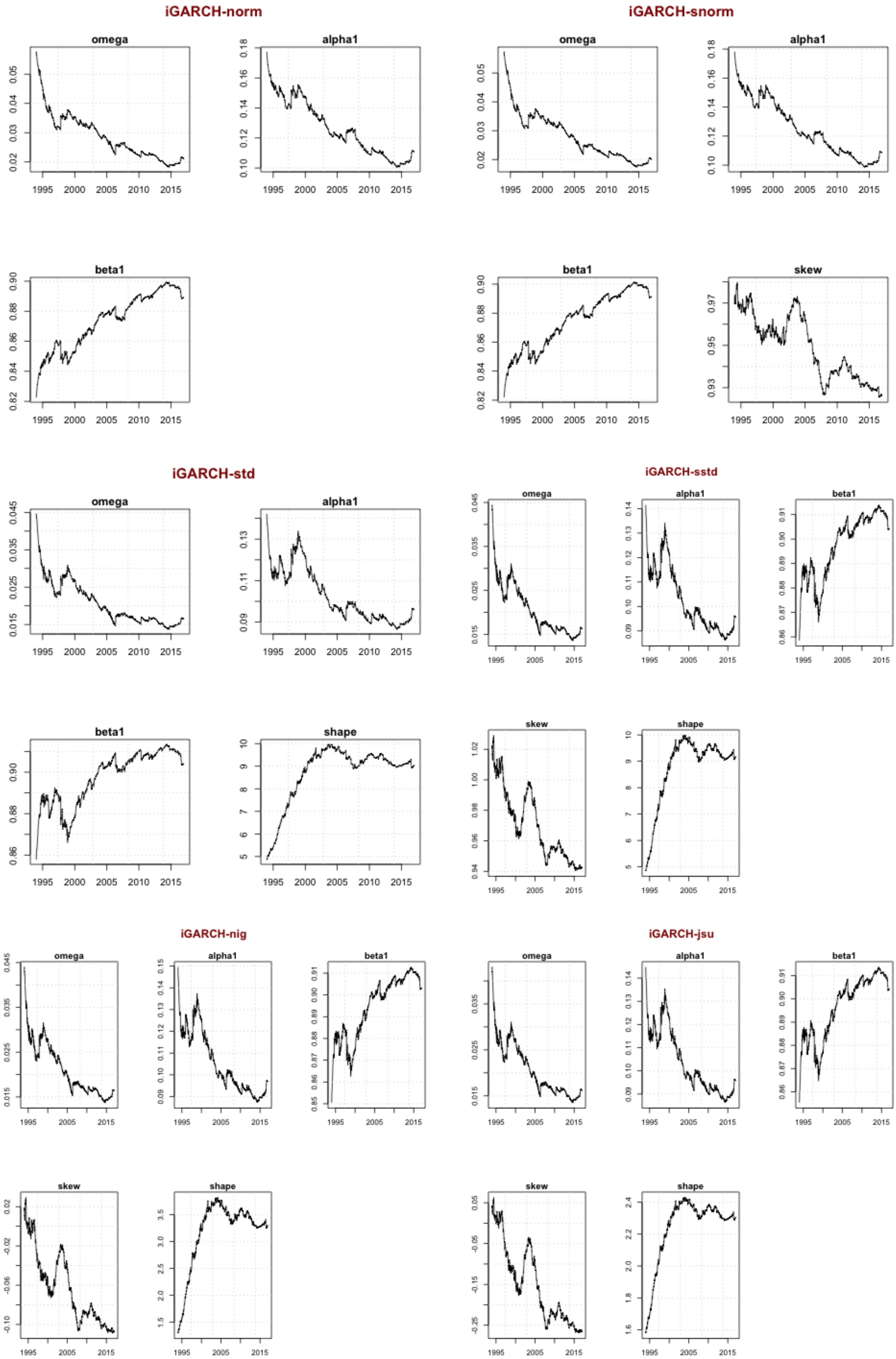

Figure B.3 Parameter plots for all the estimations of the IGARCH models.

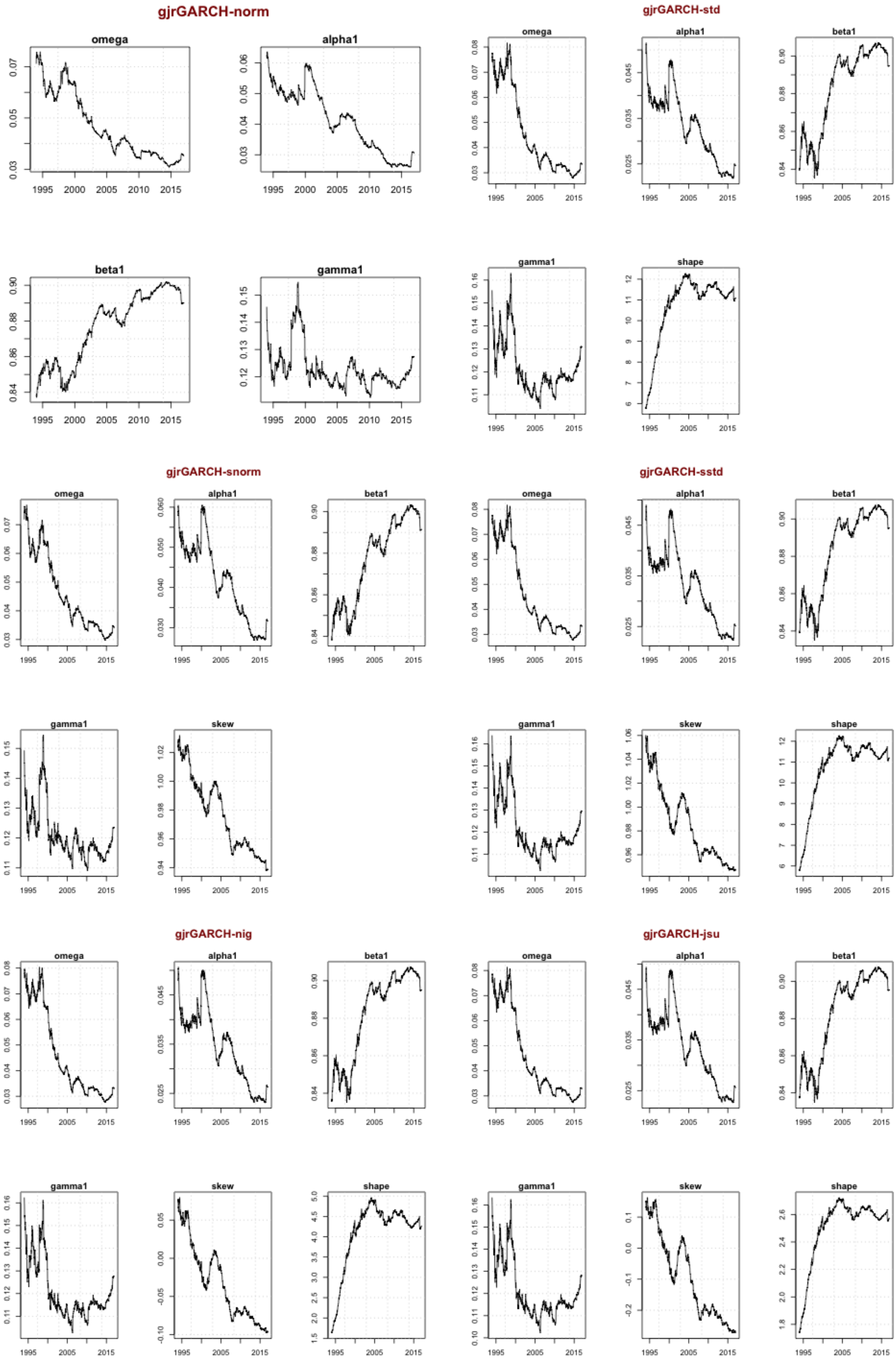

Figure B.4 Parameter plots for all the estimations of the GJR-GARCH models.

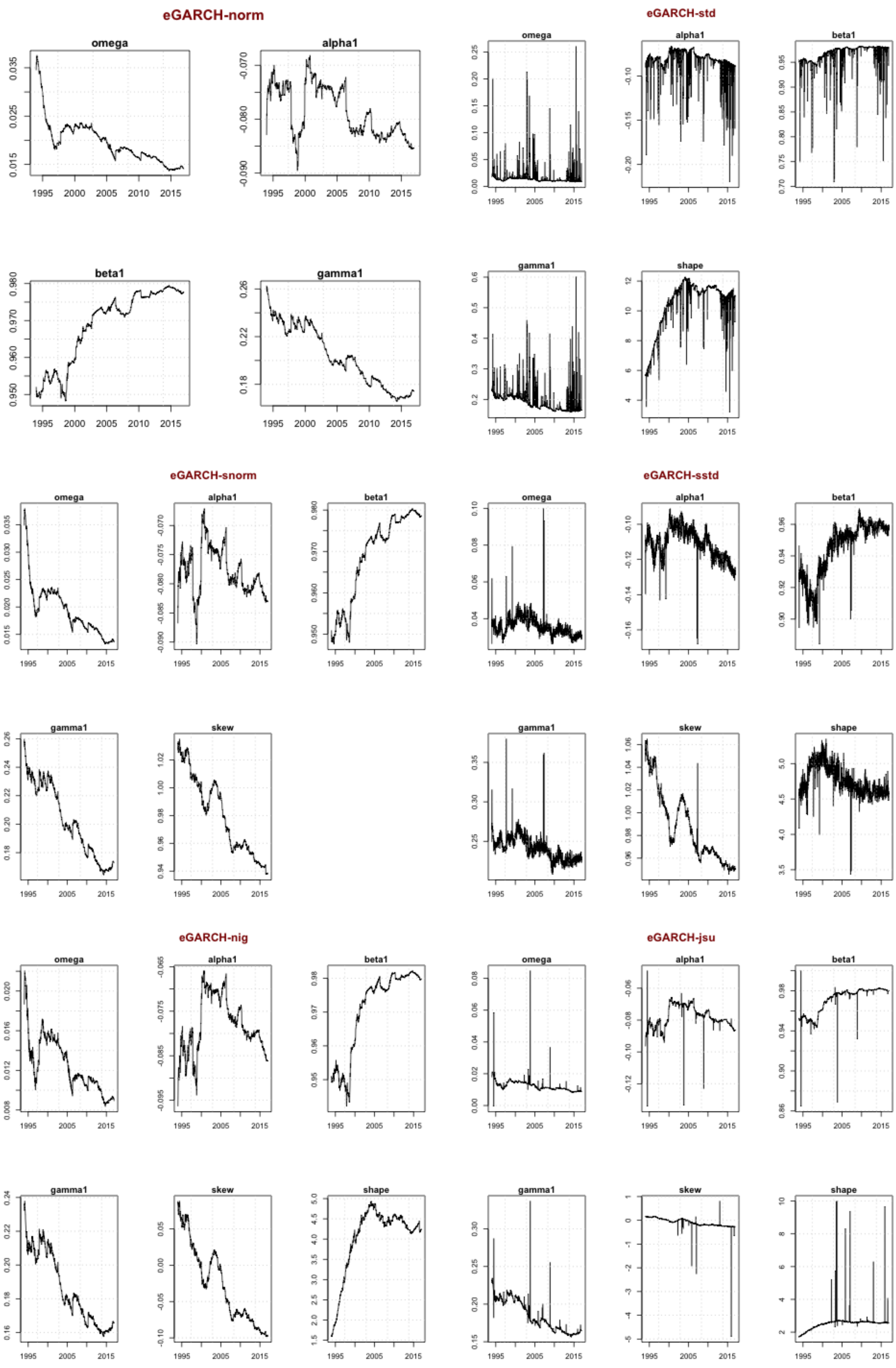

Figure B.5 Parameter plots for all the estimations of the EGARCH models.

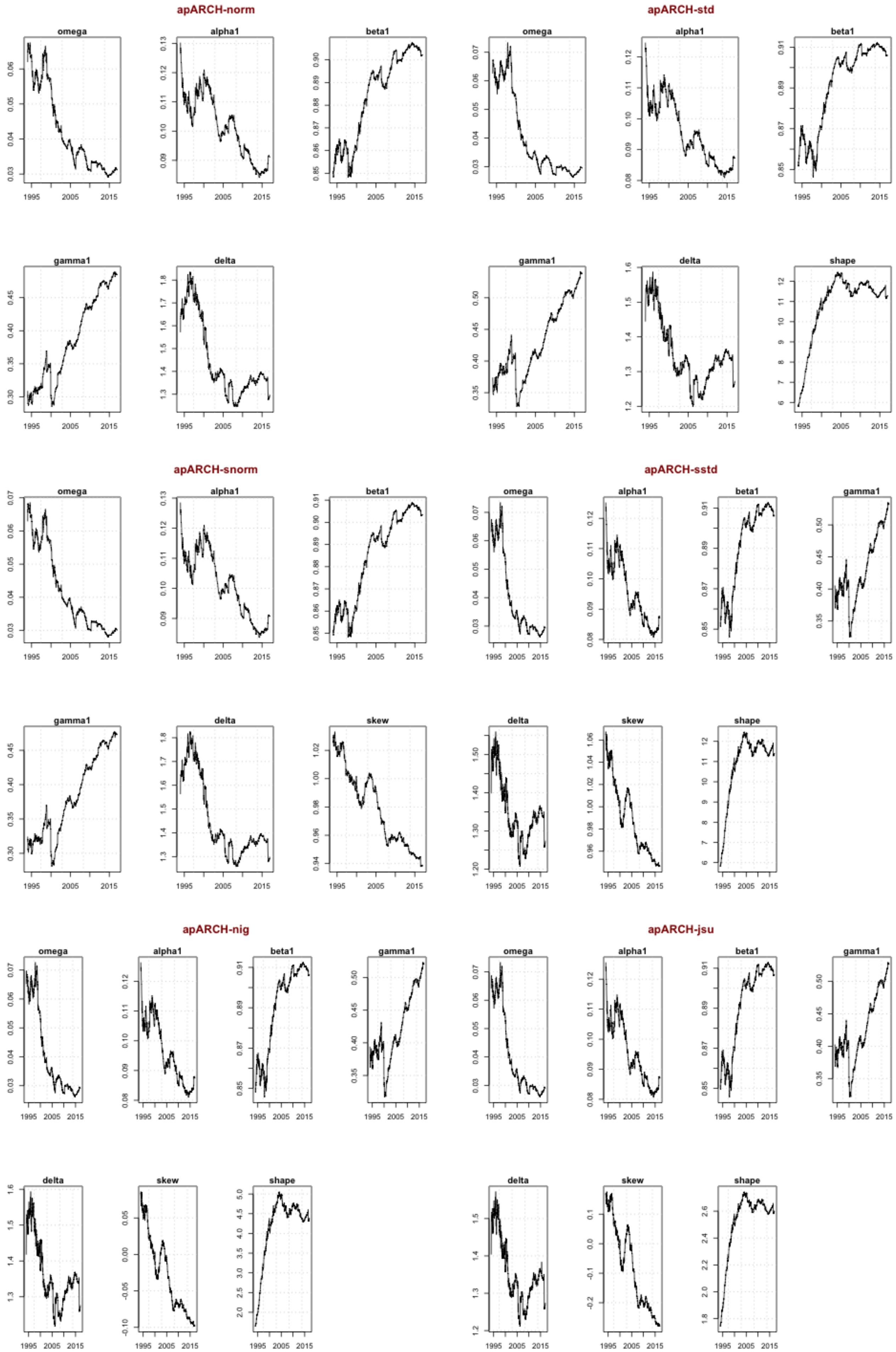

Figure B.6 Parameter plots for all the estimations of the APARCH models.

Appendix C VaR plots

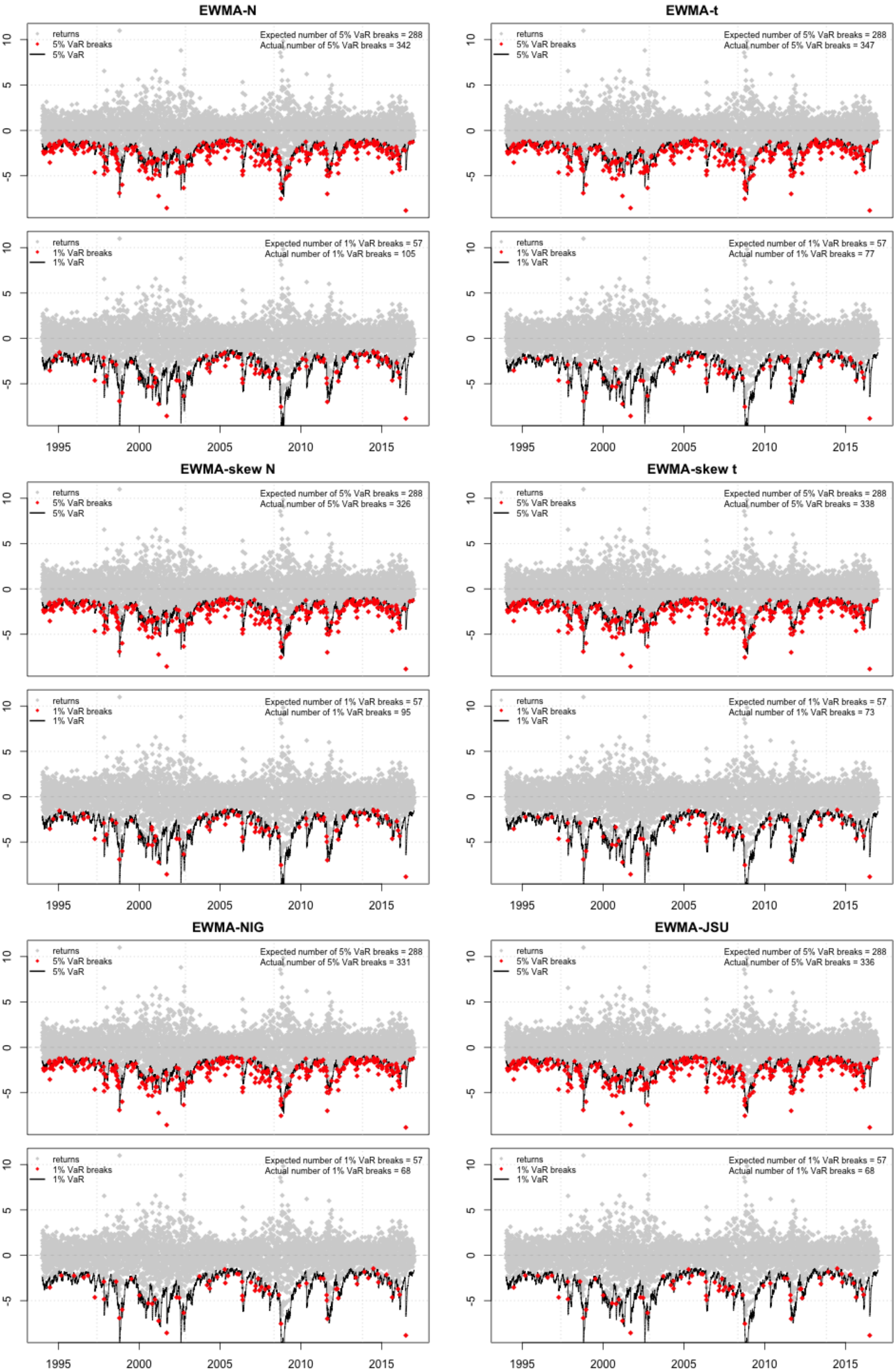

Figure C.1 Plot of OMXS30 daily returns and $\text{VaR}^{5\%}$ (top) and $\text{VaR}^{1\%}$ (bottom) forecasted one day ahead with EWMA.

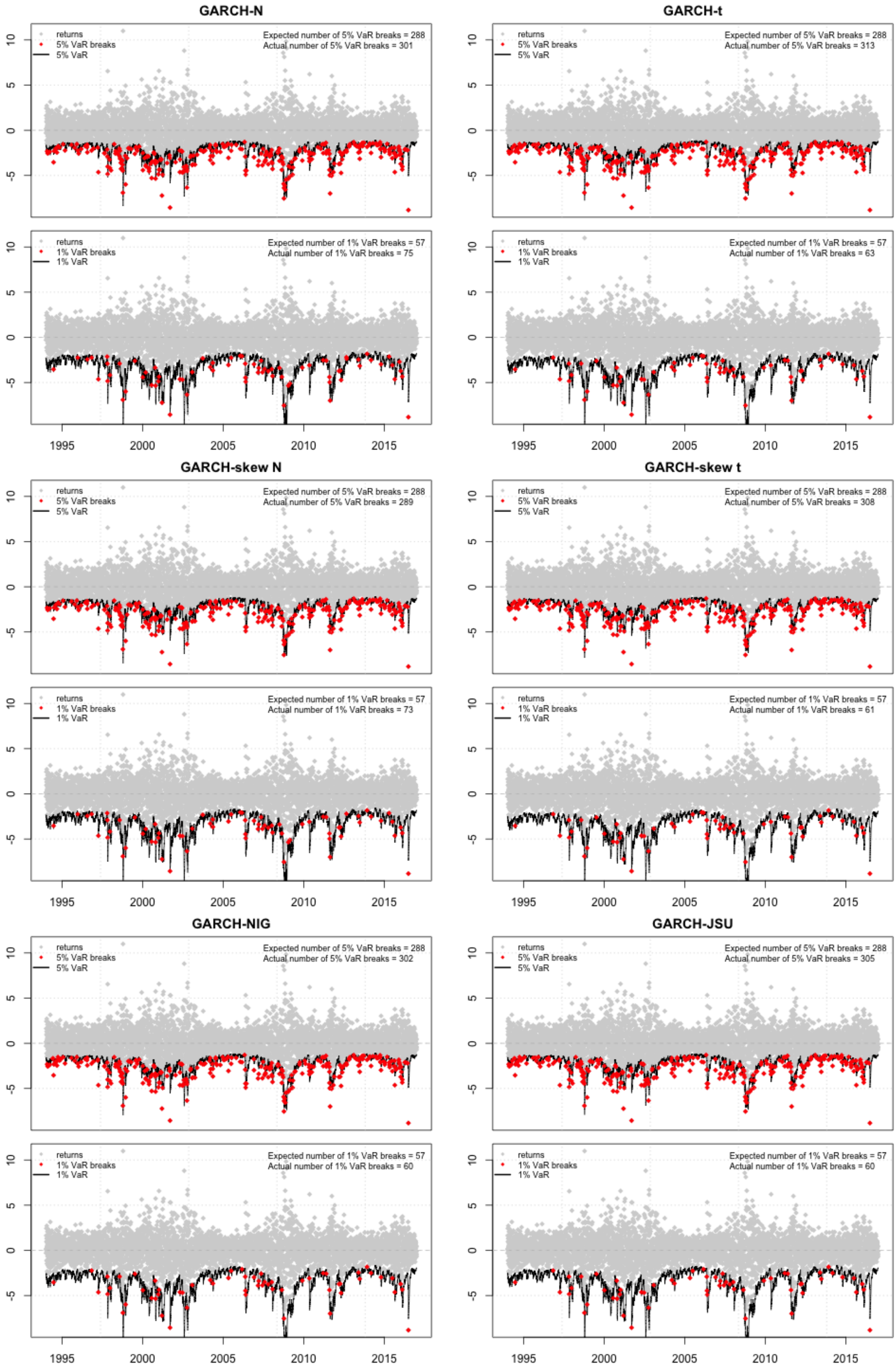

Figure C.2 Plot of OMXS30 daily returns and $\text{VaR}^{5\%}$ (top) and $\text{VaR}^{1\%}$ (bottom) forecasted one day ahead with GARCH.

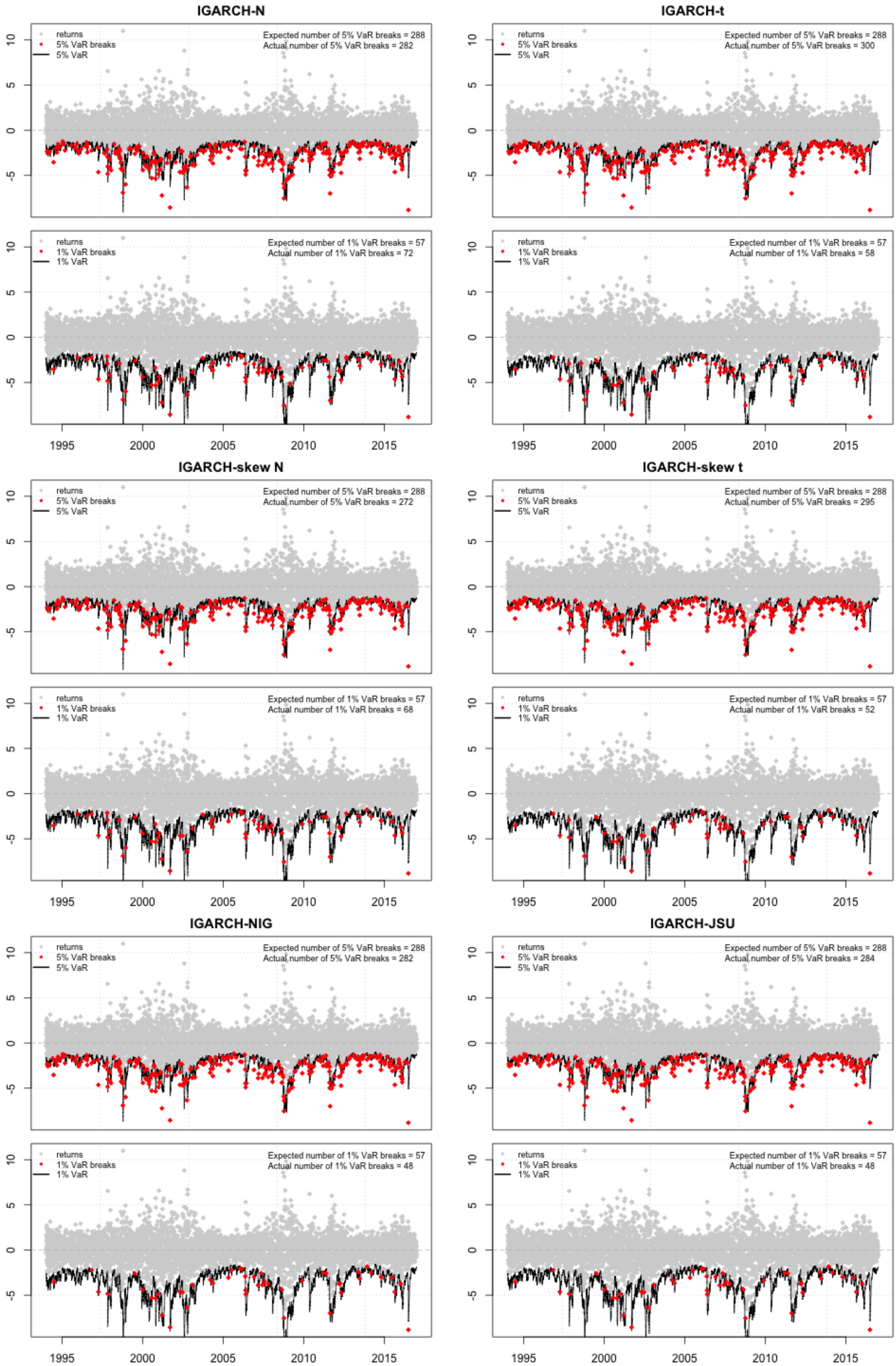

Figure C.3 Plot of OMXS30 daily returns and $\mathrm{VaR}^{5\%}$ (top) and $\mathrm{VaR}^{1\%}$ (bottom) forecasted one day ahead with IGARCH.

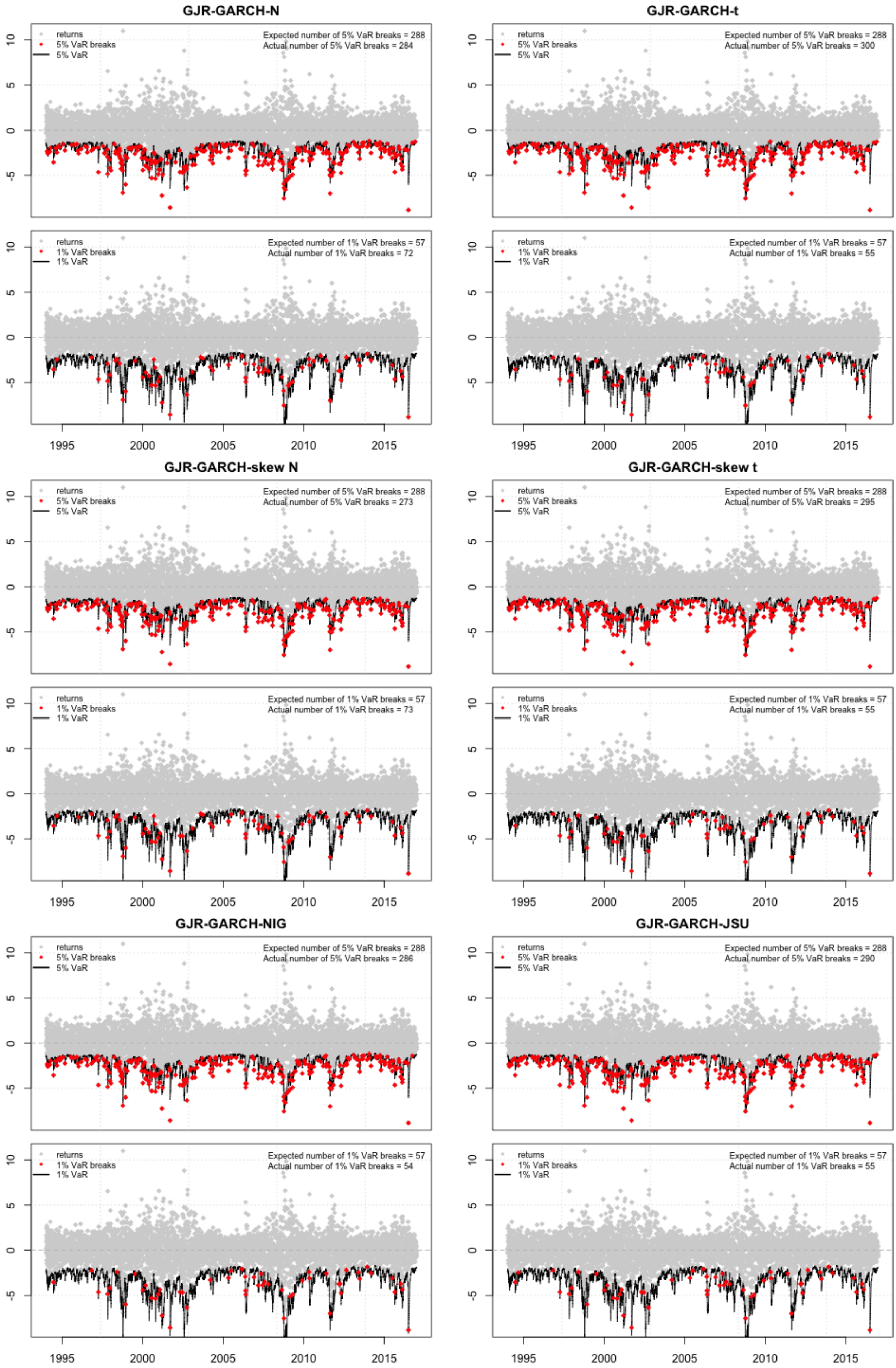

Figure C.4 Plot of OMXS30 daily returns and $\text{VaR}^{5\%}$ (top) and $\text{VaR}^{1\%}$ (bottom) forecasted one day ahead with GJR-GARCH.

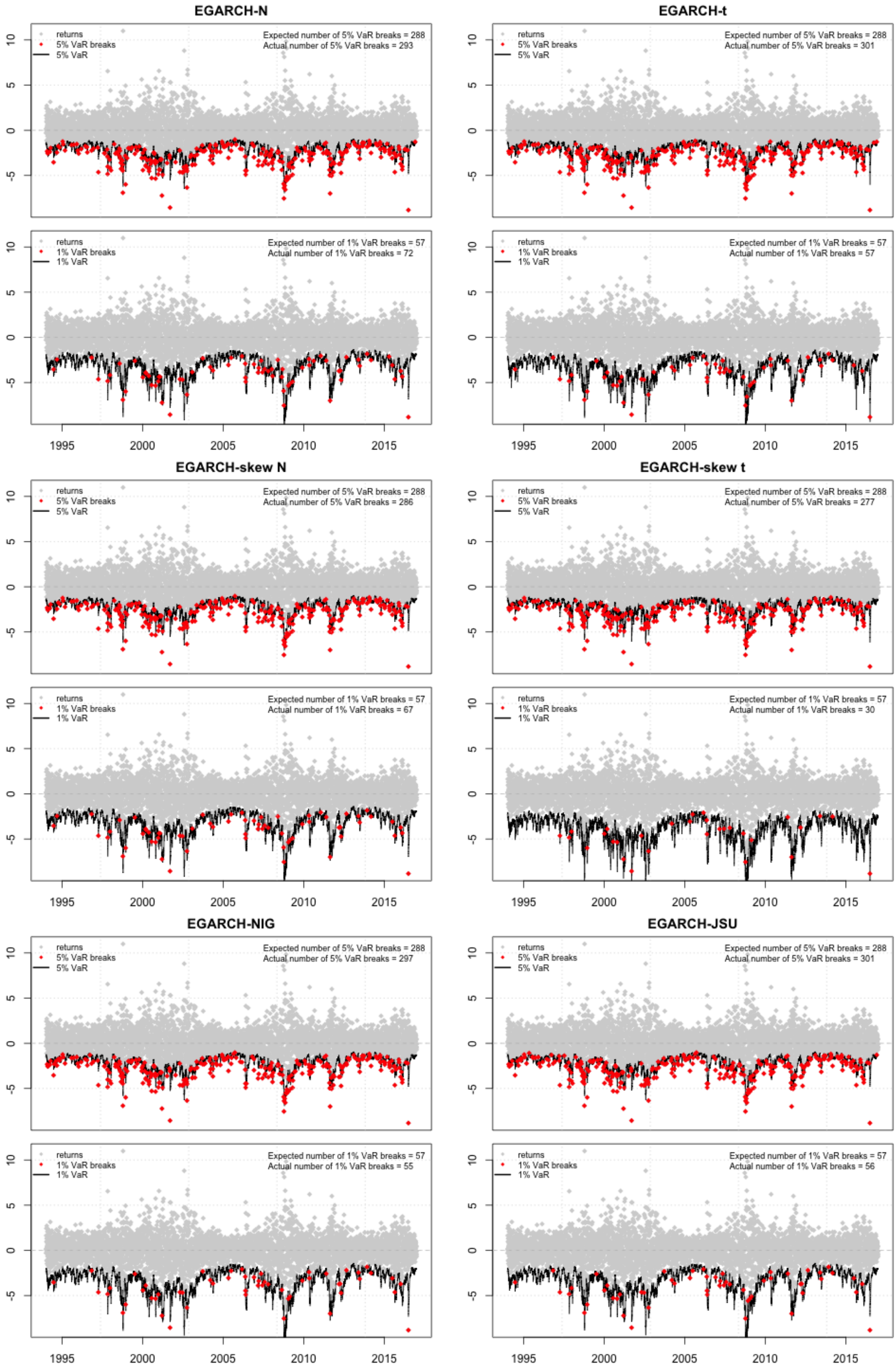

Figure C.5 Plot of OMXS30 daily returns and $\text{VaR}^{5\%}$ (top) and $\text{VaR}^{1\%}$ (bottom) forecasted one day ahead with EGARCH.

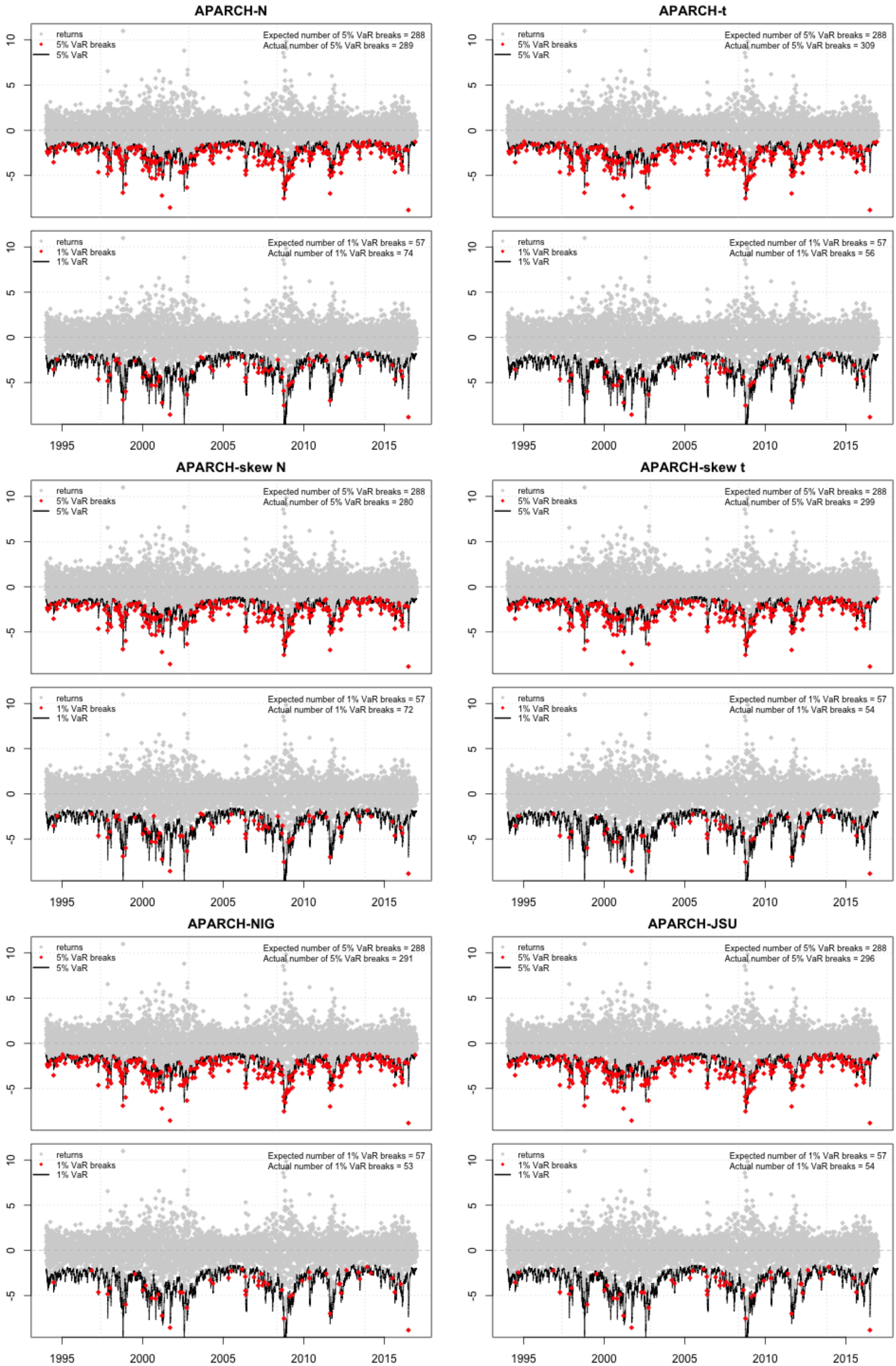

Figure C.6 Plot of OMXS30 daily returns and $\text{VaR}^{5\%}$ (top) and $\text{VaR}^{1\%}$ (bottom) forecasted one day ahead with APARCH.