Real-time GARCH: Modelling and forecasting the tail behaviour of asset returns*

Ekaterina Smetanina
University of Cambridge
Email: es599@cam.ac.uk.

Abstract

Most GARCH-type models follow Engle's (1982) original idea of modelling the volatility of asset returns as a function of only past information. We propose a new model, which retains the simple GARCH structure, but describes the volatility process as a mixture of past and current information. We show how the new model can be interpreted as the special case of a Stochastic Volatility (SV) model, which provides therefore a link between GARCH and SV models. We show that we are able to obtain better volatility forecasts than the standard GARCH-type models; improve the empirical fit to the data, especially in the tails of the distribution; and make the model faster in its adjustment to the new unconditional level of volatility. Further, we offer a much needed framework for specification testing as the new model nests the standard GARCH models.

JEL classification: C22, C51, C53

Keywords: GARCH models, High-frequency data, Forecasting

*I thank the editor, Federico M. Bandi, the associate editor and two anonymous referees for valuable comments and suggestions. I am indebted to my supervisor, Oliver Linton, for his constant encouragement and help. I would also like to thank Tim Bollerslev, Andrew Patton, Anders Rahbek, Jason Lu, Alexei Onatski, Andrew Harvey, Peter Reinhard Hansen, Torben Andersen, Stephen Thiele, Christian Pigorsch, Dario Bonciani and Ryoko Ito for useful comments. All remaining errors are mine.

1 Introduction

Volatility is widely used as a proxy for the risk associated with a financial asset, see e.g. French et al. (1987). Reliable estimation and forecasting of volatility is therefore crucial for many financial activities, such as risk management, portfolio choice and asset pricing. There are several main approaches to modelling the volatility of discrete financial time series: GARCH models (Engle, 1982; Bollerslev, 1986; Ding et al, 1993; Hansen et al., 2011, among others), Stochastic Volatility (SV) models (see Shephard, 2008 for overview), and hybrid models, e.g. Meddahi and Renault (2004). A main conceptual difference between the above approaches stems from the information structure they incorporate. Univariate GARCH models assume that the volatility of asset returns, σ_t, is a function of past information only, i.e. σ_t is $\mathcal{F}_{t-1}-$measurable, where $\mathcal{F}_{t-1}$ is the sigma-algebra induced by the history of returns up to time $t-1$. SV models assume that σ_t is $\mathcal{G}_t-$measurable, where $\mathcal{G}_t$ is the sigma-algebra induced by the history of returns as well as by the history of unobserved random shocks up to time t. The difference in the incorporated information structure is also in their nature: while GARCH models incorporate only past internal information (i.e. information generated only within the model itself) and are therefore deterministic, SV generate a stochastic volatility process by allowing for external information in the form of unobserved random shocks that are independent from the shocks governing the returns process. As a result, SV models can be more flexible in fitting the data, however this comes at the cost of higher complexity involved in their estimation and inference. Contrasting with SV models, GARCH models are observation-driven. Hence they come with the advantage of having available many estimation methods, Quasi-Maximum Likelihood (QML) being the most popular, which accounts for their wider use among practitioners.

First remarked upon by Politis (2007), by not using all available internal information, in particular the current return, GARCH models make an inefficient use of information when forecasting the volatility of returns. An important implication of this is that GARCH models are poorly suited for situations of rapid changes in financial markets, for example when volatility changes rapidly to a new level, see e.g. Andersen et al. (2003), and Hansen et al. (2011). Until now it was assumed that all volatility models can be classified as either parameter-driven or observation-driven (see e.g. Cox, 1981 and Sheppard, 1996), with a clear separation between the two. Since most GARCH models are observation-driven, it comes with a necessary condition that the process is modelled strictly in terms of the past observed information. Hence this limitation of GARCH models was believed to be inherent and unavoidable.

In this paper we show that it is possible to efficiently utilize all available internal information in GARCH models, in particular incorporating the current return. We demonstrate that by doing so, we (i) can account for rapid changes in the unconditional level of volatility as the conditional distribution of returns has a time-varying kurtosis; (ii) outperform standard GARCH models in terms of both short-run (1 and 5 days ahead) and long-run (10 and 15 days ahead) out-of-sample volatility forecasts; (iii) provide a better empirical fit to the data, especially in the tails of the distribution; (iv) provide a conceptual link between SV and GARCH models; and (v) offer a much needed framework for specification testing of the standard GARCH models, which are nested in our framework.

To put things into context, consider the following model

$$r_t = \epsilon_t \lambda_t, \quad \lambda_t^2 \text{ is } \mathcal{F}_t - \text{measurable}, \tag{1}$$

where r_t is the return series, ϵ_t are i.i.d. random variables such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = 1$, and $\mathcal{F}_t$ is the information set available at time t. Here we model the volatility as a mixture of past as well as current information, i.e. λ_t is $\mathcal{F}_t-$measurable. Compared to GARCH models, we use all information up to time t instead of time $t-1$. Compared to SV models, $\mathcal{F}_t$ contains only one source of randomness shared by the returns and volatility processes, which will allow us to retain a QML framework. The new model therefore can be thought of as a link between GARCH and SV models, as it nests the GARCH model as its special case, yet models the volatility process in the spirit of SV models where the two sources of randomness are perfectly correlated. While our model combines the advantages of both GARCH and SV models in a unified framework, it is not strictly a GARCH nor a SV model, but rather it is in a new class of its own. We call this new model the "Real-time GARCH" model (RT-GARCH for short), indicating the fact that the most "current" information is contributing to the volatility process.

An important advantage of this framework is that we allow the shape of the conditional distribution of returns to be time-varying. This has two main implications. Firstly, unlike GARCH models where the conditional kurtosis of the error terms simply translates into the kurtosis of the returns, our model's conditional kurtosis is time-varying. Secondly, the conditional density of returns is no longer a scaled normal density even when the error term has a Gaussian density. Our density function has an extra shape parameter which determines the "peakedness", and/or thickness of the tails, of the returns distribution. This allows our model to be better capture tail behaviour of the returns. This shall play an important role for the precision of our out-of-sample Value-at-risk (VaR) and short- and long-run volatility forecasts.

Politis (2007) makes the first investigation of the implications of information loss for forecasting volatility. He develops a novel model-free normalizing and variance stabilizing (NoVaS) transformation of the initial time series of returns, by incorporating the current squared returns into the conditional variance process in order to improve volatility forecasts. Being a model-free specification, parameter estimates and statistical properties are not available. Thus direct comparison of the theoretical implications of this specification with existing discrete-time volatility models is not possible, and the important question of whether including current information in a more structured model would provide any improvements over the standard GARCH models was not addressed. We answer this question by studying the statistical properties and the empirical performance of the RT-

GARCH model. We first show that it is possible to incorporate current information into GARCH-type models while retaining interpretation, and a good description of the key characteristics of financial data. We show that the new information, i.e. the current realization of the current return (or some function of thereof), can be viewed in two ways: as a change in the information set, and as providing the conditional density of returns with an extra shape parameter, making it therefore time-varying.

In our empirical study, we estimate our model on three datasets: IBM, GE and S&P 500 daily returns which span from the 2nd of January 1998 (28th of January 2003 for S&P500) till the 1st of December 2016. We find that accounting for current information in the volatility process plays an important role along several dimensions. Firstly, the RT-GARCH model outperforms standard GARCH-type models in terms of producing better short-run (1 and 5 day ahead) and especially long-run (10 and 15 days ahead) out-of-sample volatility forecasts. In particular, we compare 1-, 5-, 10- and 15-step ahead volatility forecasts with those of the GARCH(1,1) and GARCH(1,2) with standard normal and Student-t errors, APARCH(2,2) with Student-t errors, as well as NoVaS methodologies of Politis (2007). To evaluate the competing forecasts, we perform Hansen's (2011) Model Confidence Set (MCS) test and provide evidence that the RT-GARCH models always lie in the MCS for all horizons, while standard GARCH models are only occasionally included in the MCS for some datasets and/or loss functions. In particular, the MCS always contains the RT-GARCH model, and only for some datasets, the APARCH model with Student-t innovations. Moreover, the baseline RT-GARCH model always outperforms the standard GARCH(1,1) model for all horizons across all datasets. Hansen's (2005) Test for Superior Predictive Ability (SPA) confirms these results by showing that the RT-GARCH model (or variation of thereof) is not outperformed by any of the competing models. We also perform an evaluation of the forecasting performance of all models on 2 different subsamples: pre- and post-crisis periods. We show that during the crisis period, the RT-GARCH with leverage and the RT-GARCH with leverage and feedback models outperform all other models for all stocks and all horizons. This result emphasizes that during turmoil times, accounting for leverage and especially allowing for a time-varying kurtosis is crucial for getting precise forecasts. Further, using VaR as an alternative risk measurement loss function, we show that our model has the correct conditional and unconditional coverage when compared to the other models, and especially when compared to the standard GARCH(1,1) model. Secondly, being a generalization of the standard GARCH(1,1) model, the RT-GARCH model provides a better fit to the data when compared to the standard GARCH(1,1) model along several important dimensions. In particular this is most evident in the tails of the standardized residual density implied by the estimated model. Lastly, we show how the RT-GARCH model can be used for specification testing of the standard GARCH models. This specification test can be interpreted as a test for constant conditional kurtosis against a time-varying one. Applied to IBM, GE and S&P500 data, we find that all of them have a time-varying conditional kurtosis.

The remainder of the paper is structured as follows. In section 2 we introduce the RT-GARCH model and provide an interpretation of the model as well as its relation to GARCH and SV models. In section 3 we present the main results, including the conditional density function, and the strict and weak stationarity conditions. In section 4 we address the issue of leverage in the RT-GARCH model. Section 5 discusses some results of the estimation theory and the specification test. Section 6 shows how we use the RT-GARCH model to get l-step ahead volatility forecasts. In section 7 we provide an application to daily IBM, GE and S&P500 data. Section 8 concludes. All proofs are presented in the Appendix.

2 RT-GARCH

2.1 Interpretation and relation to GARCH models

In this section we formally introduce the RT-GARCH model. In order to analyze the role of current information for volatility modelling, we first need to define what is to be taken as "current information". Politis (2007) assumes that current information is represented by the current squared return. However, this poses a problem: if one is to forecast the future conditional variance at time $t+1$, the future return, r_{t+1}, will be required but is unobserved. One way to bypass this problem is to consider some function of the current return that won't require the knowledge of unobserved future returns when forecasting. It turns out that one possible candidate for doing so is the current return scaled by its volatility. In GARCH-type models this translates directly into the error term, ϵ_t, which generates the return process. This solves the forecasting infeasibility issue as only the second conditional moment of the error term will be required for forecasting, which is known for all t, provided the standard moment conditions on the error term. More precisely, consider the following joint process $(r_t; \lambda_t^2)$:

$$r_t = \lambda_t \epsilon_t \tag{2}$$

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi \underbrace{\frac{r_t^2}{\lambda_t^2}}_{=\epsilon_t^2}, \qquad (\alpha, \beta, \gamma, \varphi) \geq 0, \tag{3}$$

where r_t is the return series, ϵ_t are i.i.d. random variables with a density function $f_\epsilon(\cdot)$ such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = 1$. The true parameters are denoted by α_0, β_0, γ_0 and φ_0. This model nests the standard GARCH (1,1) model which can be obtained by setting $\varphi = 0$. We label the new volatility process λ_t^2 instead of σ_t^2, as eq.(3) does not correspond to the conditional variance of returns in this system of equations, i.e. $var[r_t^2|\mathcal{F}_{t-1}] \neq \lambda_t^2$ as λ_t is *not* independent of ϵ_t any longer. Note also that the choice of a particular function of ϵ_t, i.e. ϵ_t^2, is only one of many possible ones subject to the nesessary condition that $\lambda_t^2 > 0$. In particular, functions $|\epsilon_t|$, ϵ_t^4, among others, are possible. Our decision to choose a squared error term will become apparent later when we discuss the interpretation and the implications for the conditional distribution of returns.

Although not directly related to the MIDAS approach of Ghysels et al. (2005, 2006), as we use only one frequency, it shares a similar intuition in the sense of assigning different and, in our case, *time-varying* weights to returns on different days. In particular, it can be shown that eq.(3) can approximately be written in the following way:

$$\lambda_t^2 \approx \frac{\varphi r_t^2}{b_{t-1}} + \sum_{j=1}^{\infty} \left(\frac{\beta^j \varphi}{b_{t-1-j}} + \gamma \beta^{j-1} \right) r_{t-j}^2, \tag{4}$$

where $b_{t-1} = \alpha + \beta\lambda_{t-1} + \gamma r_{t-1}^2$. The derivation can be found in the Appendix. Compared to the standard GARCH models, the weights are time-varying and depend on past volatility, which can be approximately taken to be b_{t-1}. The intuition of this weighting scheme is as follows. For the current return r_t^2, the weight is inversely proportional to b_{t-1}, i.e. the weight is bigger for a smaller past return and is smaller if the past return is large. For any r_{t-j}^2, $j \geq 1$, the weight consists of two parts: the usual "GARCH weight", given by $\gamma\beta^{j-1}$, and an additional time-varying weight $(\beta^j\varphi)/b_{t-1-j}$ which assigns an extra weight if a particular realization of r_{t-j} is in the tails of the distribution.

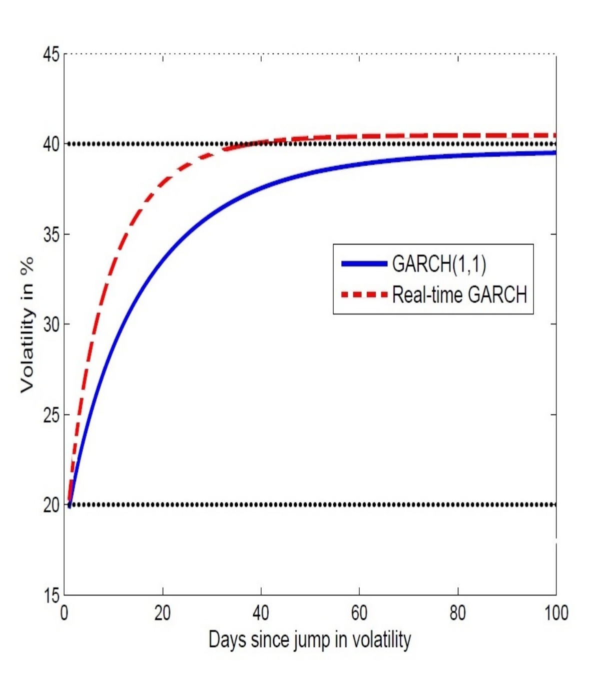
In order to understand what difference it makes to enlarge the information content of the volatility process, consider the following thought experiment which we have borrowed from the presentation of the paper by Hansen et al. (2011). Suppose that σ_t^2 is such that the volatility is $\sigma_t = 20\%$ for $t < T$, but then suddenly jumps to the new level of $\sigma_t = 40\%$ for $t \geq T$. The implication for the GARCH(1,1) model is that for any $k \geq 0$ it holds that

$$E\left(r_{T+k}^2\right) = E\left(\sigma_{T+k}^2\right) = \alpha + \gamma E\left(r_{T+k-1}^2\right) + \beta\left[\alpha + \beta E\left(\sigma_{T+k-1}^2\right) + \gamma E\left(r_{T+k-1}^2\right)\right] = \cdots =$$
$$= \frac{\alpha}{1-\beta} + \alpha \sum_{j=0}^{\infty} \beta^j E\left(r_{T+k-1-j}^2\right) = \frac{\alpha}{1-\beta} + \gamma \frac{1-\beta^k}{1-\beta}(40\%)^2 + \gamma \frac{\beta^k}{1-\beta}(20\%)^2.$$

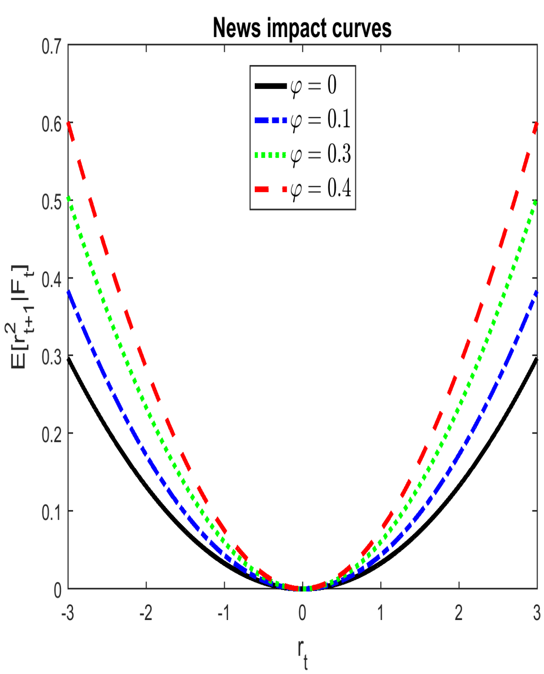
Using similar derivation steps for the RT-GARCH model with the important exception that $E\left(r_t^2\right) = E\left(\lambda_t^2\right) + \varphi\eta$, $\eta = E\left(\epsilon_t^4\right) - 1$ we similarly have:

$$E(r_{t+k}^2) = \frac{\alpha + \varphi(3-2\beta)}{1-\beta} + \gamma \frac{1-\beta^k}{1-\beta}(40\%)^2 + \gamma \frac{\beta^k}{1-\beta}(20\%)^2,$$

where we took $\epsilon_t \sim \mathcal{N}(0,1)$. In this thought experiment we ask the following question: how many days following the jump will it take for the volatility process to adjust to its new level? The answer is presented in Figure 1(a).



(a)



(b)

Figure 1: (a) Time scale of the volatility adjustment. For both graphs the parameter vector $[\ \alpha,\ \ \beta,\ \ \gamma,\ \ \varphi]$ is set to $[\ 0,\ \ 0.92,\ \ 0.073,\ \ 0.035\ ]$ for the RT-GARCH model, while $[\alpha, \beta, \gamma] = [0, 0.95, 0.045]$ for the standard GARCH(1,1) model. (b) News impact curves for different values of φ. For both graphs parameter values are in line with the ones from the estimated daily GE stock returns.

For the standard GARCH(1,1) model it takes approximately 100 days (more than 3 months) to approach the level of 39%. For the RT-GARCH model it takes a little less than 40 days to adjust to the new level of volatility. Although still slow the RT-GARCH model is at least two times faster in its speed of adjustment to the new level of volatility after a sudden jump when compared to the standard GARCH(1,1) model.

Another measure of how new information affects the volatility of returns is given by the "news impact curve", as defined by Engle and Ng (1993). For the RT-GARCH model the news impact curve is given by the following equation:

$$E\left[r_{t+1}^2|\mathcal{F}_t\right] = \alpha + \varphi\kappa + \beta\left(\frac{\bar{b} + \sqrt{\bar{b}^2 + 4\varphi r_t^2}}{2}\right) + \gamma r_t^2, \tag{5}$$

with $\kappa = E\left[\epsilon_t^4\right]$ and $\bar{b} = (\alpha + \beta\varphi + \kappa\gamma\varphi)/(1 - (\beta + \gamma))$ being the unconditional level of $b_{t-1} = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2$. Note that this news impact curve is no longer simply a quadratic function of r as in the case of standard GARCH(1,1) model. However, for reasonable values of the parameter values the last term in eq.(5) dominates. In Figure 1(b), we compare news impact curves of the RT-GARCH model for different values of φ with the news impact curve of the standard GARCH(1,1) model, which corresponds to the case of $\varphi = 0$. For a fixed value of φ the volatility in the RT-GARCH model responds much more to extreme news when compared to the standard GARCH(1,1) model. For larger values of φ this response becomes even larger, see eq.(4) for the weighting interpretation. In our baseline model good and bad news have the same weighting. We address the leverage and feedback issue and how it can be incorporated in the baseline model in Section 4.

2.2 Relation to SV models

To see how our model relates to SV models we write the simplest possible RT-GARCH and SV models, which is enough to demonstrate the point. Consider the following:

$$\left.\begin{aligned} r_t &= \lambda_t \epsilon_t \\ \lambda_t^2 &= \alpha + \varphi \epsilon_t^2, \\ \epsilon_t &\sim iid(0,1) \end{aligned}\right\} \quad RT-GARCH \qquad \left.\begin{aligned} r_t &= \sigma_t \eta_t \\ \sigma_{t+1}^2 &= w + \gamma z_{t+1}, \\ z_t \sim iid(0,\sigma_z^2), \eta_t &\sim iid(0,\sigma_\eta^2), \\ corr(z_{t+1},\eta_t) &= \rho \ \forall \ t \end{aligned}\right\} \quad \text{SV}$$

After simplifying both models as above, the difference becomes immediately clear. SV models assume that the process for returns, r_t, is driven by two random shocks, z_t and η_t. A non-zero contemporaneous dependence between shocks is allowed, which is thought to pick up the leverage effect, see Yu (2005) for the definition of the leverage effect in SV models. Note that the inter-temporal dependence between shocks can also be allowed, see also Yu (2005) for a discussion, however this can lead to returns that are not martingale difference sequences and therefore not consistent with the efficient market hypothesis. The RT-GARCH model assumes that ϵ_t, a single random shock, is common to both r_t and its volatility process λ_t. Our model is therefore a special case with $\rho = 1$ as the correlation of the shocks in the SV framework. This common shock only contributes to the volatility whenever it is large in absolute value. One therefore can think about it as really "bad" (in terms of both magnitude and sign) news that will be immediately incorporated in the volatility process. As mentioned before however, the RT-GARCH model is neither a GARCH nor a SV model, but something in between. To formally define where in-between our model lies, one would need to derive the continuous-time limit, which is currently left for future research.

3 Main Results

In this section we derive some statistical properties of the new model. We start with the unconditional moments of r_t^2 and λ_t^2. From eq.(2)-(3) the unconditional expectations of r_t^2 and λ_t^2 are given by:

$$E[r_t^2] = \alpha + \beta E[\lambda_{t-1}^2] + \gamma E[r_{t-1}^2] + \varphi E[\epsilon_t^4]$$

and

$$E[\lambda_t^2] = E[\alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi\epsilon_t^2] = \alpha + \beta E[\lambda_{t-1}^2] + \gamma E[r_{t-1}^2] + \varphi. \tag{6}$$

This now gives us a link between the first moments of λ_t^2 and r_t^2:

$$E[r_t^2] = E[\lambda_t^2] + \varphi(E[\epsilon_t^4] - 1). \tag{7}$$

When, for instance, ϵ_t are i.i.d. $\mathcal{N}(0,1)$ random variables, the above relationship simply becomes $E[r_t^2] = E[\lambda_t^2] + 2\varphi$. We next derive the conditional density of returns together with the general formula for the jth conditional moment, followed by a discussion of strict and weak stationarity conditions for r_t^2 and λ_t^2. All proofs for this section's results can be found in the Appendix.

Theorem 1. *Let ϵ_t be i.i.d. symmetric around zero random variables such that $E(\epsilon_t) = 0$ and $var(\epsilon_t) = 1$; and let (r_t, λ_t^2) evolve according to eq.(2)-(3). Denote by $\mathcal{F}_{t-1} := \sigma(r_t, s \leq t-1)$ the σ-algebra induced by the history of returns up to time $t-1$. Denote the parameter vector by $\theta = (\alpha, \beta, \gamma, \varphi)'$ and the true parameter vector by $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \varphi_0)'$. Then the conditional probability density function of the return series, $f_r(r)$, is given by*

$$f_r(r|\mathcal{F}_{t-1}) = \frac{r}{d(r, b_{t-1}, \theta)\sqrt{b_{t-1}^2 + 4r^2\varphi}} f_\epsilon(d(r, b_{t-1}, \theta)), \tag{8}$$

where $f_\epsilon(\cdot)$ is the probability density function of ϵ_t, while $d(r, b_{t-1}, \theta)$ and b_{t-1} are given by the following equations

$$d(r, b_{t-1}, \theta) = \begin{cases} \sqrt{\frac{\sqrt{b_{t-1}^2 + 4r^2\varphi} - b_{t-1}}{2\varphi}}, & for \quad \varphi \neq 0 \\ r/\sqrt{b_{t-1}}, & for \quad \varphi = 0 \end{cases} \tag{9}$$

with $b_{t-1} = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2$. Note that $\epsilon_t = d(r_t; b_{t-1}, \theta_0)$. Moreover,

$$\lim_{r\to 0} \frac{r}{d(r, b_{t-1}, \theta)} = \sqrt{b_{t-1}} \quad and \quad \lim_{r\to\ 0} f_r(r|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{b_{t-1}}} f_\epsilon(0).$$

The conditional cumulative distribution function of returns is given by:

$$F(r|\mathcal{F}_{t-1}) = F_\epsilon\left(d(r, b_{t-1}, \theta)\right),$$

where $F_\epsilon(\cdot)$ is the cdf of ϵ_t. The conditional j^{th} moment of returns, where $j \in \mathbb{Z}$, is given by the following formula:

$$E\left[r_t^j|\mathcal{F}_{t-1}\right] = b_{t-1}^{j/2}\left[E\left(d(r, b_{t-1}, \theta)^j\right) + \frac{1}{2}\frac{j\varphi}{b_{t-1}} E\left(d(r, b_{t-1}, \theta)^{j+2}\right)\right]. \tag{10}$$

Remark 1. From eq. (10) it can be noticed that for returns to be a martingale difference sequence, it is required that the third moment of ϵ_t is also zero (hence our assumption on the symmetry of the error term in the Theorems). Although definitely a stronger requirement than just $E(\epsilon_t) = 0$, we believe it is still a realistic assumption as it will hold for a variety of distributions for ϵ_t. For instance, this requirement does not rule out the densities that are multimodal as long as they are still symmetric. In particular, it will hold for the commonly used Gaussian or Student-t distributions for ϵ_t.

Remark 2. Note that the distribution related the conditional density in eq.(8) has now a time-varying kurtosis. Therefore, the parameter φ can be thought of as an extra shape parameter, representing the thickness of the tails. As a special case this distribution nests the standard Normal distribution with a constant kurtosis of 3.

Remark 3. Conditional on $\mathcal{F}_{t-1}$, r_t is an odd function of ϵ_t, since ϵ_t is an odd function of ϵ_t and λ_t is an even function of ϵ_t. It then automatically follows that the conditional, and hence unconditional distribution of r_t, is symmetric.

Remark 4. The conditional density of the RT-GARCH model in eq.(8) nests the conditional density of the standard GARCH(1,1) model as its limiting case at $r = 0$. The intuition is as follows: standard GARCH(1,1) model is a special case of the RT-GARCH model whenever $\varphi = 0$, then $d(r)$ simplifies to $r/\sqrt{b_{t-1}}$ and eq. (8) boils down to the standard GARCH(1,1) density, or $\epsilon_t = 0$, which is equivalent to the condition of $r_t = 0$. In this case the limit of eq.(8) as $r \to 0$ is again the standard GARCH(1,1) density. Similarly, the conditional moments in eq.(10) nest the GARCH(1,1) model's conditional moments as its special case.

Remark 5. It is also interesting to note another important difference with the GARCH(1,1) model for conditional moments of order $j > 2$. Recall that for the standard GARCH(1,1) model the standardized conditional kurtosis of returns is just

$$E\left[r_t^4|\mathcal{F}_{t-1}\right] / \left(E\left[r_t^2|\mathcal{F}_{t-1}\right]\right)^2 = b_{t-1}^2 E\left[\epsilon_t^4\right] / b_{t-1}^2 = E\left[\epsilon_t^4\right],$$

meaning it is simply the standardized kurtosis of the error term, and therefore constant over time. For the RT-GARCH model we have

$$E\left[r_t^4|\mathcal{F}_{t-1}\right] / \left(E\left[r_t^2|\mathcal{F}_{t-1}\right]\right)^2 = \frac{b_{t-1}^2 E\left[\epsilon_t^4\right] + 2\varphi b_{t-1} E\left[\epsilon_t^6\right] + \varphi^2 E\left[\epsilon_t^8\right]}{\left(b_{t-1} + \varphi E\left[\epsilon_t^4\right]\right)^2},$$

which makes it now time-varying. This explains why we opted to call φ an additional shape parameter, as it has a direct relationship to the standardized conditional kurtosis of the returns. In section 5 we discuss how this can be used for specification testing. Further note that the conditional distribution of the return series is no longer just the scaled version of the standard normal density. In particular, it now has an extra shape parameter φ, which, as we will describe below, will determine the degree of peakedness and/ or thickness of the tails of the distribution. More precisely, the return process described by the RT-GARCH with normal innovations is now able to account for heavier tails compared to the standard normal distribution. To highlight this point even further, Figure 2 displays the probability density function of the RT-GARCH with $f_\epsilon(\cdot) \sim \mathcal{N}(0,1)$ against the p.d.f. of the standard normal distribution and demonstrates that the density of the general model is also able to model heavier (than the standard normal) tails of the distribution without resorting to an arbitrary distribution of the error term.

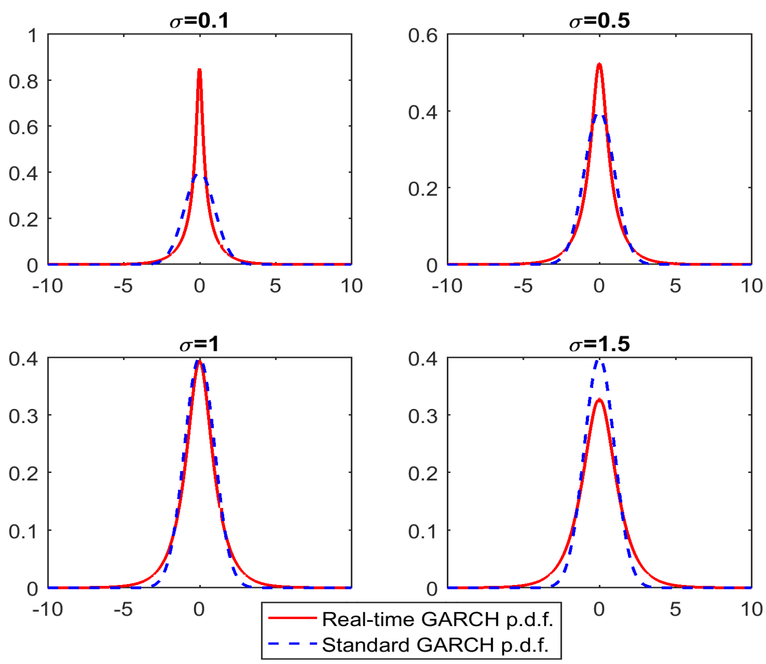


Figure 2: **Conditional probability density function for different values of unconditional volatility.** The parameter vector $\theta = [\ \alpha, \quad \beta, \quad \gamma, \quad \varphi]'$ is set to $[\ 0.003, \quad 0.9, \quad 0.04, \quad 0.02\ ]'$, which are typical parameter values.

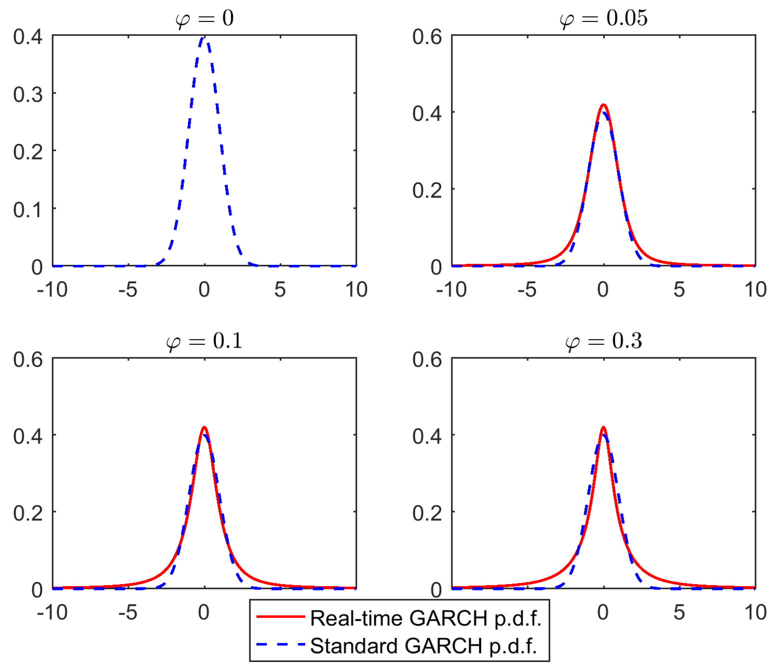


Figure 3: **Conditional probability density function for different values of φ.** The parameter vector $\theta = [\ \alpha,\ \ \beta,\ \ \gamma,\ \ \varphi]'$ is set to $[\ 0.003,\ \ 0.9,\ \ 0.04,\ \ 0.02\ ]'$, which are typical parameter values.

The reason as to why the RT-GARCH model is able to reproduce heavy tails of the returns conditional distribution stems from the fact that the additional parameter φ controls for the thickness of the tails of the corresponding distribution as the conditional kurtosis in the new model is time-varying. From the Figure 3 we can see that the larger the value of φ is the heavier are the tails of the distribution. Besides controlling for the thickness of the tails of the distribution, parameter φ allows for the adjustment of the volatility estimate, either up or down depending on the "sign" of the news, allowing the conditional variance process to be more dynamic.

After some preliminary graphical results we now turn to describing some further statistical properties of the RT-GARCH model. In particular, we derive conditions for the joint process (r_t, λ_t^2) to be strictly stationary. Establishing this result is important for developing estimation theory for the QMLE. In particular, establishing the strict stationarity conditions is important for proving that the joint process (r_t^2, λ_t^2) is geometrically ergodic and β-mixing (for proofs of these results see Smetanina (2017b)), which in turn is necessary for establishing the asymptotic normality of the QMLE.

Theorem 2. *Let ϵ_t be i.i.d. symmetric around zero random variables such that $E(\epsilon_t) = 0$ and $var(\epsilon_t) = 1$; and let (r_t, λ_t^2) evolve according to eq.(2)-(3). Let $\alpha, \beta, \gamma > 0$, and $\varphi \neq 0$. If the following conditions are satisfied*

$$-\infty \leqq E \log \left|\beta + \gamma\epsilon_0^2\right| < 0 \qquad E\left(\log\left|\alpha + \varphi\epsilon_0^2\right|\right)^+ < \infty, \tag{11}$$

then the process (r_t, λ_t^2) is strictly stationary.

We next establish the weak stationarity conditions for r_t^2 and λ_t^2 processes. These results will be later used to derive the forecasting formulae for the conditional variance of returns. In addition, the unconditional level of volatility is needed if one chooses to use variance targeting for the estimation of the parameter vector.

Theorem 3. *Let ϵ_t be i.i.d. symmetric around zero random variables such that $E(\epsilon_t) = 0$ and $var(\epsilon_t) = 1$; and let (r_t, λ_t^2) evolve according to eq.(2)-(3). Then under the following conditions:*

$$\begin{cases} \beta + \gamma < 1 & \text{(case} \quad 1) \\ \alpha + \varphi + \gamma\varphi\left(E[\epsilon_t^4] - 1\right) > 0 \end{cases}$$

or

$$\begin{cases} \beta + \gamma > 1 & \text{(case} \quad 2) \\ \alpha + \varphi + \gamma\varphi\left(E[\epsilon_t^4] - 1\right) < 0 \end{cases}$$

the process λ_t^2 is weakly stationary and its first unconditional moment is given by

$$E[\lambda_1^2] = \frac{\alpha + \varphi + \gamma\varphi\left(E[\epsilon_t^4] - 1\right)}{1 - (\beta + \gamma)}. \tag{12}$$

Given the relationship, described in eq.(7), between $E[r_t^2]$ and $E[\lambda_t^2]$, we can write down the conditions for weak stationarity of r_t^2.

Theorem 4. *Let ϵ_t be i.i.d. symmetric around zero random variables such that $E(\epsilon_t) = 0$ and $var(\epsilon_t) = 1$; and let (r_t, λ_t^2) evolve according to eq.(2)-(3). Then under the following conditions:*

$$\begin{cases} \beta + \gamma < 1 & \text{(case 3)} \\ \alpha + \varphi E(\epsilon_t^4) + \varphi\beta\,(1 - E(\epsilon_t^4)) > 0 \end{cases}$$

or

$$\begin{cases} \beta + \gamma > 1 & \text{(case 4)} \\ \alpha + \varphi E(\epsilon_t^4) + \varphi\beta\,(1 - E(\epsilon_t^4)) < 0 \end{cases}$$

r_t is weakly stationary and its second unconditional moment is given by

$$E[r_1^2] = \frac{\alpha + \varphi E(\epsilon_t^4) + \varphi\beta\,(1 - E(\epsilon_t^4))}{1 - (\beta + \gamma)}. \tag{13}$$

In addition it also holds that:

$$cov(r_t, r_s) = 0, \quad t \neq s.$$

Let us now turn to the unconditional fourth moment of the return series r_t, $E[r_1^4]$, which is an important measure of the tail behaviour of the return distribution. Detailed derivations are presented in the Appendix, here we present the resulting expression for $E[r_1^4]$.

Theorem 5. *If the process (r_t, λ_t^2) evolves according to eq. (2)-(3) and ϵ_t are symmetric around zero i.i.d. random variables such that $E(\epsilon_t) = 0$ and $var(\epsilon_t) = 1$, then r_t is fourth moment stationary if*

$$\gamma^2 < \frac{1}{E[\epsilon_t^4]}, \tag{14}$$

with the unconditional fourth moment given by

$$E[r_1^4] = \frac{\xi_1 + E[\lambda_1^2]\xi_2 + 2\beta\gamma\mu_4[E[\lambda_1^2]]^2}{1 - \gamma^2\mu_4},$$

where $\mu_j := \varphi E[\epsilon_t^j]$ and constants ξ_1 and ξ_2 are given by $\xi_1 = \alpha^2\mu_4 + \mu_8 + 2\alpha\mu_6 + 4\varphi\gamma(\alpha\mu_4 + \beta\mu_4 + \mu_6) > 0$ and $\xi_2 = \mu_4(2\alpha\beta + \beta^2 + 2\alpha\gamma + 2\mu_6(\gamma + \beta)) > 0$ and $E[\lambda_1^2]$ is given by eq. (12).

Remark 6. In the case of Gaussian error terms, condition (14) simply becomes $\gamma^2 < \frac{1}{3}$ which is exactly the same as in the standard GARCH(1,1) case.

4 Leverage and volatility feedback effects

The RT-GARCH model described by eq. (2)-(3) has no leverage effect, meaning that when errors are symmetric about zero, $E(r_t) = 0$ and $cov(r_t^2, r_j) = 0 \ \ \forall j$. However, there is well documented empirical evidence, see e.g. Black (1976), Christie (1982), Engle and Ng (1993), that many financial time series exhibit the leverage effect, i.e. the contribution to the volatility of negative shocks to the stock prices is far greater than that of the positive shocks of the same magnitude. As a result of this empirical evidence, most discrete and continuous-time volatility models were extended to incorporate this feature. For discrete time models see Nelson (1991), Engle and Ng (1993), Glosten et al.(1993) among others. For continuous-time models, see Christie (1982), Yu (2005), Bandi and Renò (2012), Aït-Sahalia et al. (2013) and Wang and Mykland (2014). For fully nonparametric way of estimating and testing the leverage hypothesis, see also recent work by Linton et al. (2016).

We proceed by incorporating the leverage effect in the fashion of Glosten et al. (1993), i.e. by acknowledging the different effect of positive and negative news on the conditional variance of returns. Note however that, unlike for all standard GARCH-type models, the most recent information in our case is represented by current shocks ϵ_t. We therefore refer to "leverage effect by differentiating the effect of positive and negative values of ϵ_t on λ_t^2. Therefore the baseline model in Section 2 can be extended to account for leverage effect as follows:

$$r_t = \lambda_t \epsilon_t$$

and

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi_1 \epsilon_t^2 \mathbb{1}_{(\epsilon_t > 0)} + \varphi_2 \epsilon_t^2 \mathbb{1}_{(\epsilon_t \leq 0)}.$$

It is also interesting to differentiate between the effect of positive and negative values of past returns on the conditional volatility. In the standard GARCH-type models this is referred as "leverage effect" as this would be the most recent information effecting the conditional volatility. Given the differently defined leverage effect in our model we refer to the different effects of the past postive and negative returns on conditional variance as "feedback effect". More precisely, the RT-GARCH model with leverage and feedback effects is given by

$$r_t = \lambda_t \epsilon_t \tag{15}$$

and

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma_1 r_{t-1}^2 \mathbb{1}_{(r_t > 0)} + \gamma_2 r_{t-1}^2 \mathbb{1}_{(r_t \leq 0)} + \varphi_1 \epsilon_t^2 \mathbb{1}_{(\epsilon_t > 0)} + \varphi_2 \epsilon_t^2 \mathbb{1}_{(\epsilon_t \leq 0)}. \tag{16}$$

In Figure 4 we compare the news impact curves of the GJR-GARCH(1,1) model of Glosten et al.(1993) with the RT-GARCH model with leverage and the RT-GARCH model with leverage and feedback, all estimated on the daily IBM data. For both specifications of the RT-GARCH model, volatility tends to respond more to negative news than in GJR-GARCH model. Interestingly, the RT-GARCH model with leverage and feedback responds slower to negative news than the RT-GARCH just with the leverage effect.

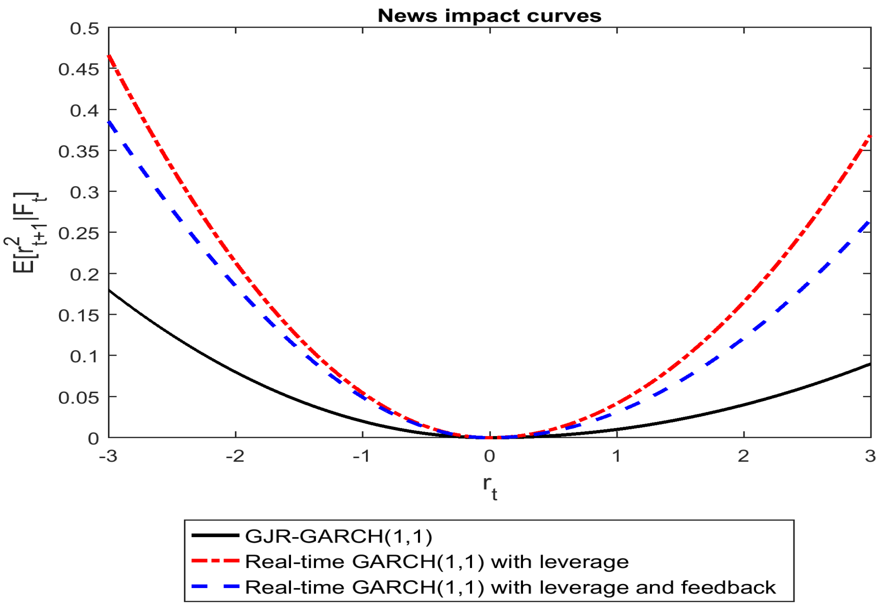


Figure 4: The figure displays the news impact curves for three models, estimated on the daily IBM data.

All theorems in section 3 hold for both extensions with slight modifications. For reasons of brevity we defer these to the Supplementary Material for this paper.

5 Outline of the Estimation Theory

In this section we discuss some results of the QMLE analysis. We denote by $\theta = (\alpha, \beta, \gamma, \varphi)'$ the parameter vector and the corresponding true parameter vector by $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \varphi_0)'$. For the purpose of estimation we adopt a Gaussian specification, such that the log-likelihood function can be written as follows:

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta),$$

where

$$l_t(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}d_t^2(r, b_{t-1}, \theta) + \log\left(\frac{\sqrt{b_{t-1}(\theta) + \varphi d_t^2(r, b_{t-1}, \theta)}}{b_{t-1}(\theta) + 2\varphi d_t^2(r, b_{t-1}, \theta)}\right),$$

with $b_{t-1}(\theta) = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2$, and $d_t^2(r, b_{t-1}, \theta)$ is given in eq.(9). Note also that if φ is set to zero we are again back to the standard GARCH(1,1) log-likelihood function. The QMLE of θ is then defined as any measurable solution $\widehat{\theta}_T$ of $\widehat{\theta}_T = \arg\max_{\theta\in\Theta} L_T(\theta)$, where Θ denotes the parameter space. Given that the RT-GARCH(1,1) model nests the standard GARCH(1,1) model, it can be expected that the asymptotics theory for QMLE will be a generalization of some sort for the standard GARCH(1,1) model. In fact, this turns out to be true, however the analysis is non-trivial and requires a lengthy derivations. In addition, we believe that the entire analysis is beyond the scope of this paper as it presents an interest of its own. The details therefore can be found in Smetanina (2017b), and here we provide only a brief discussion of the results. Importantly, the joint process (r_t^2, λ_t^2) remains to be a Markov chain and therefore the theory for Markov models, developed by Meitz and Saikkonen (2008) applies. This allows one to establish the ergodicity and $\beta-$mixing of the process. After the dependence structure is established, in Smetanina (2017b) we show that the strong consistency of $\widehat{\theta}$ can be established by adopting the theory by Francq and Zakoïan (2004). In addition, we also show that the score function is still a martingale difference sequence, therefore the martingale CLT, see e.g. Hall and Heyde (1980), can be applied to show:

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, V_\theta\right),$$

where $V_\theta \equiv A^{-1}BA^{-1}$ and

$$A = -\frac{1}{T}E_{\theta_0}\left[\frac{\partial^2 \log L_T(\theta)}{\partial\theta\partial\theta'}\right] \quad \text{and} \quad B = \frac{1}{T}E_{\theta_0}\left[\frac{\partial \log L_T(\theta)}{\partial\theta}\frac{\partial \log L_T(\theta)}{\partial\theta'}\right].$$

The exact expressions for V_θ can be found in Smetanina (2017b). Finally, provided that $\widehat{\theta} \xrightarrow{p} \theta_0$ and $\widehat{V}_\theta \xrightarrow{p} V_\theta$, the feasible version is given by:

$$\widehat{V}_{\widehat{\theta}}^{-1/2}\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, I\right).$$

We finish this section by suggesting that the new model can be used for specification testing of the standard GARCH models. In particular, one can consider testing the following null hypothesis:

$$\mathbb{H}_0 : \quad \varphi = 0$$

versus an alternative hypothesis $\mathbb{H}_A$ that $\mathbb{H}_0$ is false. This test can be interpreted as the test for constant standardized conditional kurtosis of the returns against an alternative of a time-varying conditional kurtosis. Since this test is for nested models, it is straightforward to use already computed likelihood quantities to calculate the Likelihood Ratio test $LR = -2\ln(L_T(\theta^\star)/L_T(\theta)) \xrightarrow{d} \chi^2_1$, where $\theta = (\alpha, \beta, \gamma, \varphi)'$, $\theta^\star := \{\theta \setminus \varphi\}$. Although theoretically φ can take negative values, see Theorems 3 and 4 for restrictions, in practical applications the easiest way to ensure that λ_t^2 is always positive is to restrict all parameters to be positive, i.e. $\varphi \geq 0$. In this case the test is on the boundary of the parameter space for φ, and the Likelihood Ratio test has a nonstandard distribution, see Francq and Zakoïan (2009) for details.

6 Volatility Forecasts with RT-GARCH

We now focus on volatility forecasting using the RT-GARCH model. The forecasting exercise is very similar to obtaining volatility forecasts with the standard GARCH(1,1) model except for some slight differences. Recall that for the forecasting exercise we need the following two equations:

$$E[\lambda_t^2|\mathcal{F}_{t-1}] = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi \tag{17}$$

and

$$E[r_t^2|\mathcal{F}_{t-1}] = E[\lambda_t^2|\mathcal{F}_{t-1}] + \varphi(E[\epsilon_t^4] - 1), \tag{18}$$

where eq. (17) is the expectation of the conditional variance of the process and eq. (18) is obtained by recursively substituting eq. (3) into the squared eq. (2) and taking expectations. Then k-step ahead volatility forecast is given by:

Theorem 6. *Let the process (r_t, λ_t^2) evolve according to eq. (15)-(16) and ϵ_t is symmetric around zero i.i.d. random variables such that $E(\epsilon_t) = 0$ and $var(\epsilon_t) = 1$. Then the k-step ahead, $k \geq 1$, volatility forecast is given by the following formula:*

$$E[r_{t+k}^2|\mathcal{F}_t] = E[\widehat{\lambda}_{t+k}^2|\mathcal{F}_t]+(\widehat{\varphi}_1+\widehat{\varphi}_2)(E[\epsilon_{t+k}^4|\mathcal{F}_t]-1) = E[\lambda_1^2]+(\widehat{\beta}+\widehat{\gamma}_1+\widehat{\gamma}_2)^k\left(E[\widehat{\lambda}_t^2|\mathcal{F}_t]-E[\lambda_1^2]\right)+$$
$$+(\widehat{\varphi}_1+\widehat{\varphi}_2)(E[\epsilon_{t+k}^4|\mathcal{F}_t]-1),$$

where $E[\lambda_1^2]$ is given by

$$E\left[\lambda_1^2\right] = \frac{\alpha + (\varphi_1 + \varphi_2)\left[\eta(\gamma_1+\gamma_2)+1\right]}{1-(\beta+\gamma_1+\gamma_2)},$$

with $\eta \equiv E\left[\epsilon_t^4\right] - 1$ and $E\left[\widehat{\lambda}_t^2|\mathcal{F}_t\right]$ is an estimate of λ_t^2 at time t.

Note that Theorem 6 provides the most general formulae for the RT-GARCH with leverage and feedback. Forecasting formulae for the RT-GARCH model with leverage only may be obtained by setting $\gamma_1 = \gamma_2 = \gamma$, while for the basic RT-GARCH model by setting $\gamma_1 = \gamma_2 = \gamma$ and $\varphi_1 = \varphi_2 = \varphi$.

7 Application

7.1 Data and Methodology

To estimate and evaluate competing models we use 3 datasets of open-to-close returns, namely IBM, General Electric (GE) and the S&P 500 index. For IBM and GE the data spans from the 2nd of January 1998 till the 1st of December 2016, while for the S&P500 the time span is from the 28th of January 2003 till the 1st of December 2016. The detailed description of the data is provided in the Supplementary Material. To avoid estimation bias we split the sample into two parts, the first part will serve for the model's estimation, and the remaining part will be used for an evaluation of the out-of-sample performance using an expanding window scheme. As with any forecasting exercise there is no direct guidance of the optimal splitting point. For presenting the main results, we reserve 2/3 of the whole sample for the estimation and the rest of the sample for the forecast evaluation. For the IBM and GE stocks this results in 3000 and 1500 observations for estimation and evaluation respectively. For the S&P500 data we have 2000 and 1000 observations for estimation and evaluation respectively. However, to make sure that our results do not depend on the splitting point, we present results for different splitting points in the Supplementary Material. In addition to our full sample results, due to the likelihood of structural breaks during the financial crisis period we also present results for 2 subsamples: pre- and post-crisis periods. The pre-crisis period spans from 2nd of January (28th of January 2003 for the S&P 500 index) till the end of July 2008. The crisis and post-crisis period constitutes the rest of the available sample. For subsamples for GE and IBM stocks we take 1500 and 500 obsevations for estimation and evaluation respectively. For subsamples for S&P500 data we use 1000 and 500 obsevations for estimation and evaluation respectively.

For out-of-sample forecast performance we compare RT-GARCH models with the standard GARCH(1,1), GARCH(1,2) with normal and Student-t innovations, APARCH model with Student-t distributed innovations (the most sophisticated GARCH-type model, see Hansen and Lunde (2005) for details), as well as Simple and Exponential NoVaS methodologies of Politis (2007). The specifications of all competing models are presented in Table 1. We exclude SV models from this comparison as SV models are outside of the Maximum Likelihood framework. Moreover, since the purpose of this paper is not to propose the best volatility model but rather investigate whether incorporating available current information in GARCH-type models will improve on existing GARCH models in terms of out-of-sample volatility forecasts, inclusion of SV models is not necessary to answer this question.

The "true volatility" would be needed in order to directly evaluate the forecasting performance of competing models. Without the true volatility process, the most common approach instead is to use realized volatility as a proxy for the conditional variance of returns. We calculate the 5-minute realized variance from the intraday high-frequency data for each stock, which we then take to be the proxy for the conditional variance of returns in out-of-sample forecast evaluations.

7.2 Results and Discussion

In this section we report the parameter estimates for the RT-GARCH, the RT-GARCH with leverage effect (RT-GARCH-L) and the RT-GARCH with leverage and feedback effect (RT-GARCH-LF) models. Results are presented in Table 2. For all RT-GARCH models and all datasets the parameter φ is positive and significantly different from zero. Note that for the model with leverage, the value of the parameter φ_2 is much larger than the value of the parameter φ_1, pointing at the fact that negative news contribute to volatility more than positive ones.

For out-of-sample evaluation we use the only two "robust" loss functions (see Patton, 2011) in the context of volatility forecasting. A loss function is "robust" if for any two volatility forecasts, h_{1t}^2 and h_{2t}^2, their ranking according to expected loss is equivalent whether it is done using the true conditional variance, σ_t^2, or some proxy $\hat{\sigma}_t^2$, provided the latter is conditionally unbiased, i.e. $E[r_t^2|\mathcal{F}_{t-1}] = E\left[\hat{\sigma}_t^2|\mathcal{F}_{t-1}\right] = \sigma_t^2$.

Tables 3-14 present the results. For the presentation of results we adopt the original notation of Hansen et al. (2011), i.e. $\widehat{\mathcal{M}}^\star_{95\%}$ denotes the MCS $\widehat{\mathcal{M}}^\star$ that contains the best models with probability 0.95. For both statistical loss functions, MSE and QLIKE, Real-time-GARCH and RT-GARCH-L models are always in the MCS $\widehat{\mathcal{M}}^\star_{95\%}$ for all horizons, while standard GARCH models most of the time fall outside of the MCS. We present the results for full sample as well as the results for pre- and post-crisis (including crisis) subsamples.

We start by discussing the full sample results. For the 1-step ahead out-of-sample volatility forecasts using the MSE loss function, the MCS for the IBM stock is quite wide and consists of all competing models except for the NoVaS methodologies, while for the QLIKE loss function the MCS consists solely of all RT-GARCH models. For the GE stock for 1-step ahead forecasts MCS consists of all RT-GARCH models and the APARCH(2,2) model for both loss functions. Finally, for the S&P 500 stock the MCS based on MSE loss function is quite small and consists only of RT-GARCH and RT-GARCH-L models, while the MCS based on the QLIKE loss function consists of RT-GARCH, RT-GARCH-L and APARCH(2,2) models.

For the 5-step ahead forecasts the picture is very similar, except that for MSE loss function MCS sometimes includes the GARCH models with Student-t innovations. For example, for the 5-step ahead forecasts using IBM data for the MSE loss function the MCS consists only of RT-GARCH model, while for the QLIKE loss function both GARCH models with Student-t innovations are included as well. A similar picture can be seen for the GE stock for the MSE loss function, while for the QLIKE loss function the MCS consists again only of RT-GARCH, RT-GARCH-L and the APARCH(2,2) models. For the S&P 500 stock for the MSE loss function the MCS consists of all competing models but NoVaS methodologies, while for the QLIKE loss function the MCS consists only of the RT-GARCH and APARCH(2,2) models.

For longer horizons, i.e. 10- and 15-step ahead out-of-sample volatility forecasts, the picture is quite different. For all datasets the MCS consists only of the RT-GARCH and the APARCH(2,2) models with the occasional inclusion of RT-GARCH-L model and sometimes GARCH models with Student-t innovations.

Interestingly, for all horizons the standard GARCH models with Gaussian innovations are excluded from the MCS for all stocks. It is also interesting to note that most of the time the MCS for all datasets contains models with Student-t innovations (which allows for heavier tails) and RT-GARCH models. However, RT-GARCH models perform no worse (or most of the time even better) with just the normal innovations. As discussed in section 2 the possible reason for this is that RT-GARCH models account for a time-varying conditional kurtosis, therefore allowing the volatility to adjust to a new level faster than the other standard GARCH models. It is also possible that the forecasting performance of RT-GARCH models can be further improved if one considers Student-t innovations for the error term. On the other hand, estimated on the full sample the RT-GARCH model with leverage and feedback effects (RT-GARCH-LF) seems to perform worse than the simple RT-GARCH or RT-GARCH-L, as it can potentially overfit the data due to the model's higher complexity (i.e. higher number of parameters).

Given that all samples under consideration include the financial crisis, it is important to account for the structural break in the volatility of returns. If one is to account for the structural break, the parameters of each model have to be re-estimated during/after the break. We address this issue by estimating and evaluating the models on two subsamples: pre- and post-crisis period, where the latter includes the crisis period as well.

While the forecast evaluation results for the pre-crisis period are quite similar to the full sample results, the crisis period MCS is quite different for all stocks. For the crisis and post-crisis period the MCS for both loss function mainly consists of RT-GARCH-L, RT-GARCH-LF and the APARCH(2,2) models. This result is general for all stocks and all horizons. The difference in results emphasizes that during volatile periods it is crucial to account for both leverage and time-varying kurtosis.

There are several reasons why NoVaS methodologies are never in the MCS. First of all, Politis (2007) compares forecasts with the Mean Absolute Deviation (MAD) loss function, which as we now know, due to Patton (2011), is not a robust loss function in the context of volatility forecasting. The other reason may be that the comparison of NoVaS forecasts was done with the use of squared returns as a volatility proxy, which was shown to be quite a noisy proxy for volatility of returns, see Hansen and Lunde (2006).

We also evaluate our forecasts with the risk management loss function, i.e. we compute 1-step VaR forecasts using all competing models. For evaluation of VaR forecasts we compute the Violation Ratio (VR), which is the ratio between the number of returns that exceeded the VaR forecast to the number of the expected exceedances, accounting for a significance level of α which we take to be 5%. If the model is accurate, the violation ratio is expected to be exactly 1. A model has good forecasts if the VR is between 0.8 and 1.2; and a model has quite imprecise forecasts if the VR<0.5 or VR>1.5. However, computing only the VR is not enough for evaluating VaR forecasts as it is the measure of the unconditional coverage. We therefore also compute the Likelihood Ratio (LR) for the conditional coverage from the failure process of the VaR forecasts, see Christoffersen (1998) for details. Table 15 presents the results. Out of all models with a correct conditional coverage, RT-GARCH (for all stocks) and RT-GARCH with leverage (for IBM stock) are the only models that have an acceptable VR. In addition, this ratio will be far better than for the standard GARCH(1,1) model with normal errors for all stocks under consideration. This result further emphasizes the effect of having a time-varying kurtosis of returns, which allows for the possibility of adjusting it over time in response to the data, playing a potentially crucial role for forecasting.

After identifying which models are in the MCS, it is still interesting whether we can pin down a single superior (in the sense that it is not outperformed by any other competing model) forecasting model among those in the MCS. One possibility is to conduct an out-of-sample test that has the ability to control either for possible over-fitting or over-parametrization problems, which gives a more powerful framework to evaluate the performances of competing models. We choose to conduct Hansen's (2005) Test for Superior Predictive Ability (SPA). For reasons of brevity results of the SPA test are presented in the Supplementary Material to this paper. The overall conclusion is that the winning model (among those in the MCS) is one of the RT-GARCH models for shorter horizons (i.e. 1-

and 5-step ahead) and either APARCH model or RT-GARCH/RT-GARCH-L for longer horizons.

In addition, we perform the likelihood ratio test for $\mathbb{H}_0 : \varphi = 0$, adjusted for testing on the boundary, see Francq and Zakoïan (2009) for details. The values of the test statistic are 8.5, 4.66 and 9.72 for IBM, GE and S&P500 respectively, which are significant at a 5% significance level. This suggests that all time series have a time-varying conditional kurtosis.

Moreover, to show that the RT-GARCH model is a better fit to the data, especially in the tails, figures 5-7 display the QQ plots of the standardized errors from the estimated GARCH(1,1) and RT-GARCH models.

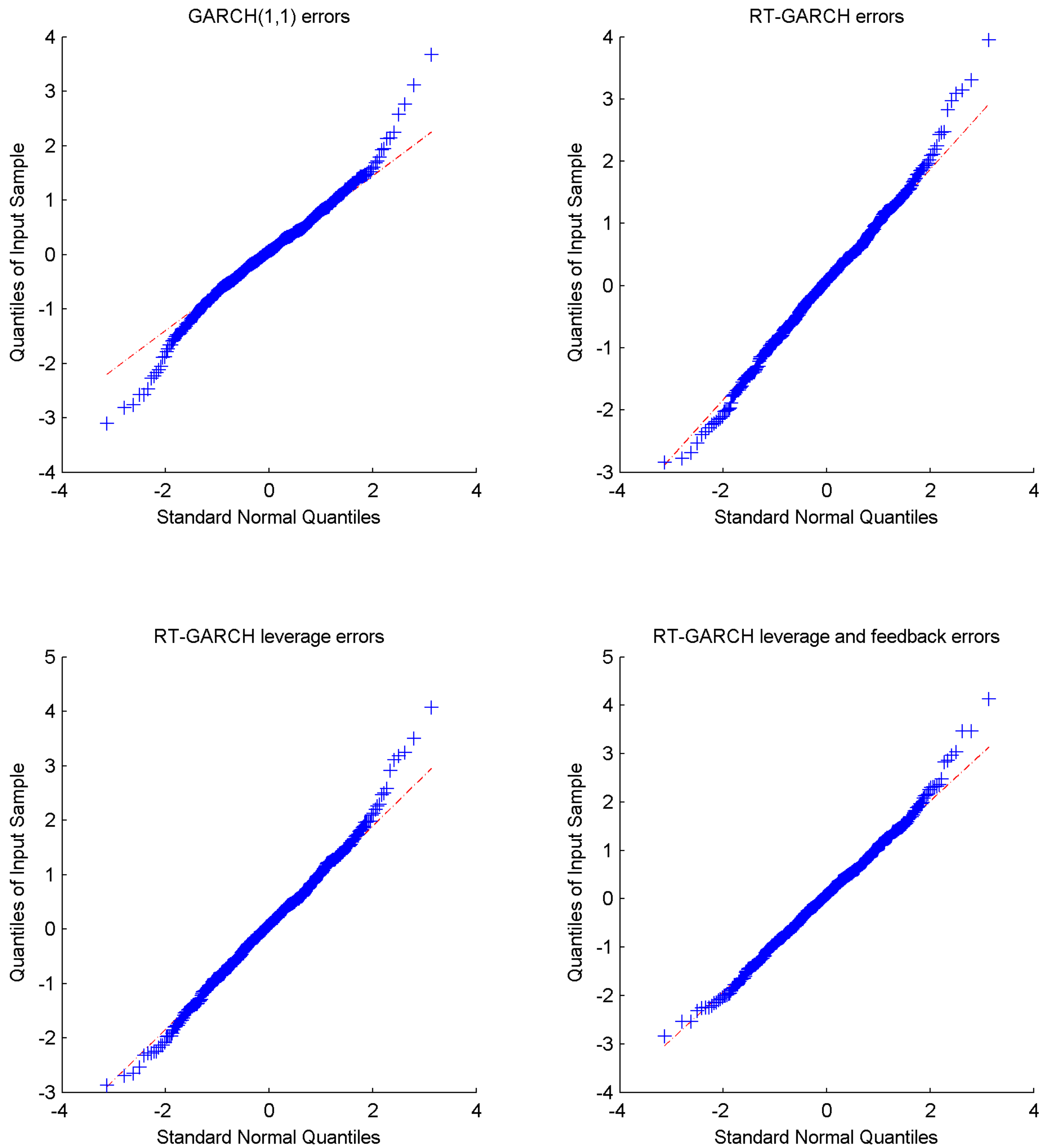


Figure 5: QQ-plots of the implied error distribution for IBM stock.

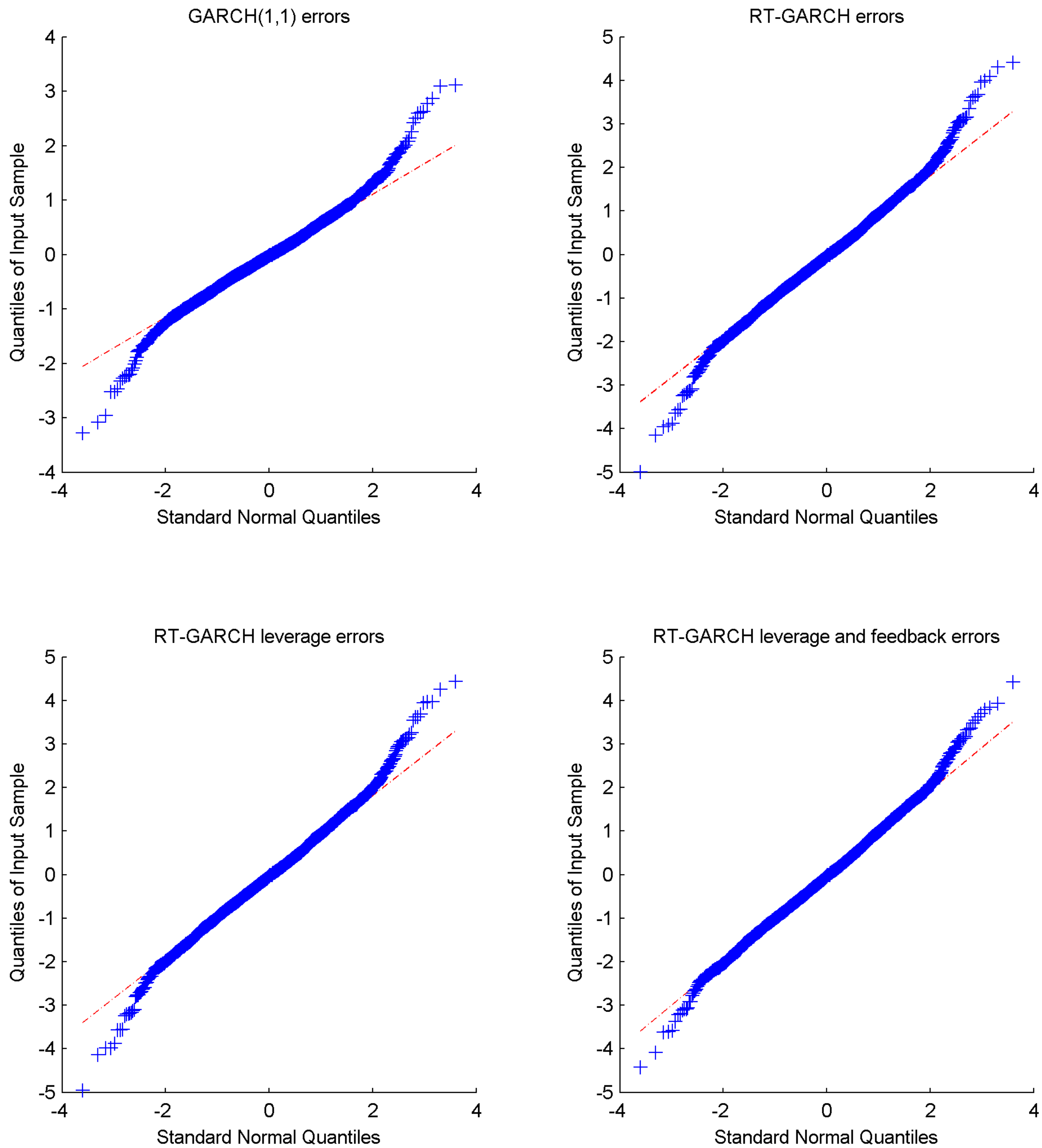


Figure 6: QQ-plots of the implied error distribution for GE stock.

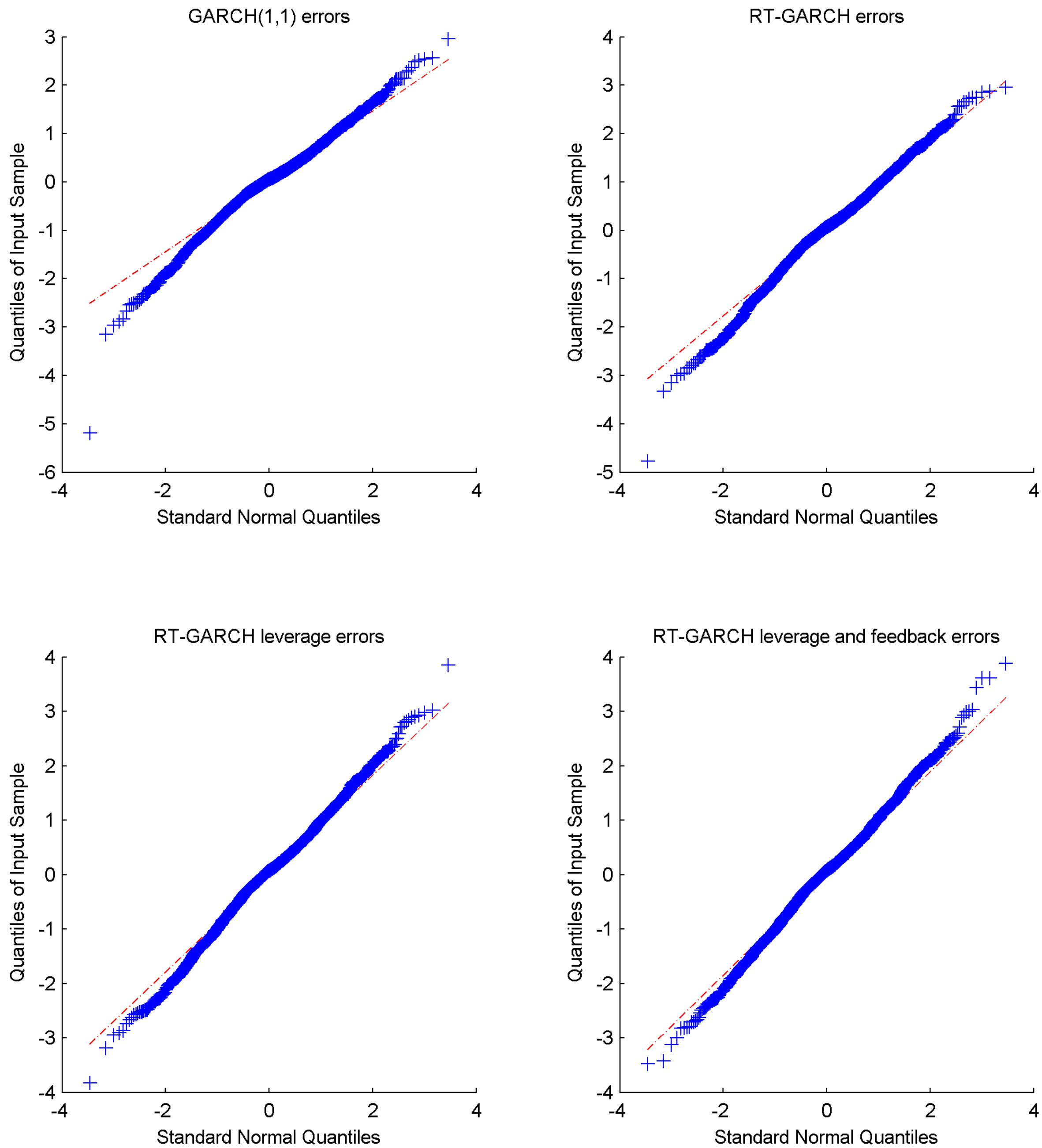


Figure 7: QQ-plots of the implied error distribution for S&P500 index.

8 Conclusion

Volatility of asset returns is difficult to forecast due to its latent nature. In an attempt to describe the volatility process standard GARCH models incorporate only past information in modelling volatility of assets' returns. Up until now there was no evidence on the relevance of incorporating current information into the conditional variance modelling in GARCH-type models. We fill this gap by proposing a new model, the RT-GARCH, which incorporates current information. The model is very general; it nests the standard GARCH models as its special case, and can easily incorporate leverage and feedback effects by differentiating between positive and negative news. The new term, i.e. the current realization of the standardized return, can be viewed in two ways: as a change in the information set, and as an extra shape parameter for the density of returns which determines the "peakedness" and/or thickness of the tails. This shape parameter allows the conditional distribution of returns to have a time-varying kurtosis, which accounting to the empirical application may well play a crucial role in forecasting volatility during turbulent times.

Estimation of the RT-GARCH revealed that (i) incorporating current information into volatility modelling allows the model to respond quicker to sudden changes of the unconditional level of volatility; and (ii) the combination of ex-ante and ex-post volatility measurement helps to improve out-of-sample volatility forecasts and empirical fit when compared to the forecasts and empirical fit given by the other competing models. Moreover the new model offers a framework for specification testing, which can be thought of a test for constant conditional kurtosis versus a time-varying one.

We finish by suggesting some routes for future research. It would be of interest to investigate whether the empirical performance of the proposed model can be further improved by incorporating some realized measures as in Hansen et al.(2011) and/or assuming Student-t distribution for innovations. In addition, deriving a continuous-time limit of the RT-GARCH model will provide an answer of where exactly between GARCH and SV models it stands. We leave these suggestions for future research.

References

[1] Aït-Sahalia, Y., J. Fan and Y. Li, 2013, The leverage effect puzzle: disentangling sources of bias at high frequency, *Journal of Financial Economics*, 109, 224-249.

[2] Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P., 2003, Modeling and forecasting realized volatility, *Econometrica*, 71, 529-626.

[3] Bandi, F. M. and R. Renò, 2012, Time-Varying leverage effects, *Journal of Econometrics*, 169, 94-113.

[4] Black, F., 1976, Studies of stock price volatility changes, in *Proceedings of the 1976 Meetings of the American Statistical Association*, 171-181.

[5] Bollerslev, T., 1986, Generalized autoregressive heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

[6] Brandt, A., 1986, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, *Advances in Applied Probability*, 18, pp. 211-220.

[7] Christie, A. A., 1982, The stochastic behavior of common stock variances: value, leverage and interest rate effects, *Journal of Financial Economics*, 10, 407-432.

[8] Christoffersen, P. F., 1998, Evaluating interval forecasts, *International Economic Review*, 1998, 39(4), 841862.

[9] Cox, D.R., 1981, Statistical analysis of time series: some recent developments, *Scandinavian Journal of Statistics*, 8(2), 93-115.

[10] Ding, Z., Granger, C.W.J. and Engle, R.F., 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, vol. (1), pp. 83-106.

[11] Engle, R. F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, Vol. 50, pp. 987-1007.

[12] Engle, R. F. and V. K. Ng, 1993, Measuring and testing the impact of wews on volatility, *The Journal of Finance*, 48, 1749-1778.

[13] Francq, C. and Zakoïan, J.-M., 2004, Maximum likelihood estimation of GARCH and ARMA-GARCH processes, *Bernoulli*, 10, 605-637.

[14] Francq, C. and Zakoïan, J.-M., 2009, Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons, *Journal of the American Statistical Association*, 104, 313-324.

[15] French, K.R., Schwert, G.W. and Stambaugh, R.F., 1987, Expected stock returns and volatility, *Journal of Financial Economics*, 19(1), 3-29.

[16] Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, There is a risk-return tradeoff after all, *Journal of Financial Economics*, 76, 509-548.

[17] Ghysels, E., P. Santa-Clara, and R. Valkanov, 2006, Predicting volatility: getting the most out of return data sampled at different frequencies, *Journal of Econometrics*, 131, 59-95.

[18] Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48(5), 1779-1801.

[19] Hall, P. and Heyde, C., 1980, Martingale limit theory and its application. New York: Academic Press.

[20] Hansen, P.R., 2005, A test for superior predictive ability, *Journal of Business&Economic statistics*,23(4), 365-380.

[21] Hansen, P.R., Huang, Z. and Shek, H.H., 2011, Realized GARCH: a joint model for returns and realized measures of volatility, *Journal of Applied Econometrics*,27, 877-906.

[22] Hansen, P.R. and Lunde, A., 2001. A forecast comparison of volatility models: does anything beat a GARCH (1,1)?, *Journal of Applied Econometrics*, vol. 20, 873-889.

[23] Hansen, P.R. and Lunde, A., 2006, Consistent ranking of volatility models, *Journal of Econometrics*, Vol. 131, 97-121.

[24] Hansen P.R., Lunde, A. and Nason J. M., 2011, Model confidence set, *Econometrica*,Vol. 79, 453-497.

[25] Kesten, H., 1973, Random difference equations and renewal theory for products of random matrices, *Acta Mathematica*, Vol. 131, Issue 1, 207-248.

[26] Linton, O., Whang Y-J. and Yen, Y-M., 2016, A nonparametric test of a strong leverage hypothesis, *forthcoming in Journal of Econometrics.*

[27] Meddahi, N. and E. Renault, 2004, Temporal aggregation of volatility models, *Journal of Econometrics,* 119, 355-379.

[28] Meitz M., and Saikkonen P., 2008, Ergodicity, Mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, *E*conometric Theory, 24, 1291-1320

[29] Nelson, D. B., 1991, Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347-370.

[30] Patton, A., 2011, Volatility forecast evaluation and comparison using imperfect volatility proxies *, Journal of Econometrics*, 160(1), 246-256.

[31] Politis, D. N., 2007, Model-free versus model-based volatility prediction, *Journal of Financial Econometrics*, Vol. 5, Issue 3, 358-359.

[32] Shephard, N., 2008, Stochastic volatility, in New Palgrave Dictionary of Economics, 2nd Edition, MacMillan.

[33] Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility, Time Series Models in Econometrics, Finance and Other Fields, London: Chapman & Hall, 1-67.

[34] Smetanina, E., 2017b, Asymptotic inference for Real-time GARCH model, *Working paper, University of Cambridge*, available at SSRN: http://ssrn.com/abstract=2809055.

[35] Stout, W.F., 1974, Almost sure convergence, probability and mathematical statistics, New York: Academic press.

[36] Wang, D. C. and P. A. Mykland, 2014, The estimation of leverage effect with high frequency data, *Journal of the American Statistical Association*, 109, 197-215.

[37] Yu, J., 2005, On leverage in a stochastic volatility model, *Journal of Econometrics*, 127, 165-178.

Table 1: The conditional variance specification of different models.

RT-GARCH	$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi\epsilon_t^2, \quad E[r_t^2] = E[\lambda_t^2] + 2\varphi$
RT-GARCH with leverage	$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi_1\epsilon_t^2\mathbb{1}_{(\epsilon_t \geq 0)} + \varphi_2\epsilon_t^2\mathbb{1}_{(\epsilon_t<0)}, \quad E[r_t^2] = E[\lambda_t^2] + 2(\varphi_1 + \varphi_2)$
RT-GARCH with leverage and feedback	$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma_1 r_{t-1}^2\mathbb{1}_{(r_t \geq 0)} + \gamma_2 r_{t-1}^2\mathbb{1}_{(r_t<0)} + \varphi_1\epsilon_t^2\mathbb{1}_{(\epsilon_t \geq 0)} + \varphi_2\epsilon_t^2\mathbb{1}_{(\epsilon_t<0)}$ $E[r_t^2] = E[\lambda_t^2] + 2(\varphi_1 + \varphi_2)$
GARCH(1,1) with standard normal errors	$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2$
GARCH(1,2) with standard normal errors	$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma_1 r_{t-1}^2 + \gamma_2 r_{t-2}^2$
GARCH(1,1) with Student's t-distributed errors	$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2$
GARCH(1,2) with Student's t-distributed errors	$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma_1 r_{t-1}^2 + \gamma_2 r_{t-2}^2$
APARCH(2,2) with Student's t-distributed errors	$\sigma_t^2 = \alpha_0 + \alpha_1\left[\lvert r_{t-1}\rvert - \gamma_1 r_{t-1}\right]^2 + \alpha_2\left[\lvert r_{t-2}\rvert - \gamma_2 r_{t-2}\right]^2 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2$
Simple NoVaS	$\sigma_t^2 = \alpha s_{t-1}^2 + \alpha_0 X_t^2 + \sum_{i=1}^{p}\alpha_i X_{t-p}^2$ $s_{t-1}^2 = \frac{1}{t-1}\sum_{k=1}^{t-1} X_k^2, \alpha = 0, \alpha_i = \frac{1}{p+1}, 0 \leq i \leq p$
Exponential NoVaS	$\sigma_t^2 = \alpha s_{t-1}^2 + \alpha_0 X_t^2 + \sum_{i=1}^{p}\alpha_i X_{t-p}^2$ $s_{t-1}^2 = \frac{1}{t-1}\sum_{k=1}^{t-1} X_k^2, \alpha = 0, \alpha_i = c'^{e^{-ci}}, 0 \leq i \leq p, c' = \frac{1}{\sum_{i=0}^{p} e^{-ci}}$

Note: In Simple NoVaS p is chosen to match the kurtosis (=3) of the normalized return series. In Exponential NoVaS initial value of p is chosen to be $\frac{n}{5}$; c is chosen to match the kurtosis (=3) of the normalized return series and p is adjusted by the maximization routine. APARCH(2,2) corresponds to the standard GJR(2,2) model whenever $0 \leq \gamma_i \leq 1$, $i = 1, 2$.

Table 2: Parameter estimates of RT-GARCH models

Parameter estimates of RT-GARCH

Dataset	α	β	γ	φ
IBM	0.0006	0.8755	0.0780	0.0758
	$(18 * 10^{-4})$	$(9 * 10^{-4})$	$(14 * 10^{-4})$	$(21 * 10^{-4})$
GE	0.0001	0.9211	0.0627	0.0378
	$(14 * 10^{-4})$	$(38 * 10^{-3})$	$(2 * 10^{-5})$	$(17 * 10^{-4})$
S&P 500	0.0001	0.9124	0.0726	0.0138
	$(12 * 10^{-4})$	$(14 * 10^{-3})$	$(45 * 10^{-3})$	$(11 * 10^{-4})$

Parameter estimates of RT-GARCH with leverage

Dataset	α	β	γ	φ_1	φ_2
IBM	0.0003	0.8883	0.0703	0.0475	0.0886
	$(15 * 10^{-4})$	$(6 * 10^{-4})$	$(11 * 10^{-4})$	$(19 * 10^{-4})$	$(27 * 10^{-4})$
GE	0.0001	0.9273	0.0550	0.0237	0.0529
	$(2.7 * 10^{-4})$	$(38 * 10^{-4})$	$(4.2 * 10^{-4})$	$(2 * 10^{-4})$	$(48 * 10^{-3})$
S&P 500	0.0016	0.8995	0.0718	0.0003	0.0481
	$(25 * 10^{-4})$	$(15 * 10^{-3})$	$(6.7 * 10^{-4})$	$(27 * 10^{-4})$	$(8.1 * 10^{-4})$

Parameter estimates of RT-GARCH with leverage and feedback

Dataset	α	β	γ_1	γ_2	φ_1	φ_2
IBM	0.0001	0.8599	0.0328	0.0706	0.0903	0.1319
	$(17 * 10^{-4})$	$(30 * 10^{-4})$	$(14 * 10^{-4})$	$(15 * 10^{-4})$	$(26 * 10^{-4})$	$(29 * 10^{-4})$
GE	0.0001	0.9225	0.0343	0.0322	0.0450	0.1253
	$(15 * 10^{-4})$	$(40 * 10^{-3})$	$(11 * 10^{-4})$	$(10 * 10^{-4})$	$(18 * 10^{-3})$	$(27 * 10^{-3})$
S&P 500	0.0023	0.9185	0.0127	0.0605	0.0004	0.0740
	$(4.2 * 10^{-4})$	$(1.8 * 10^{-3})$	$(5 * 10^{-4})$	$(23 * 10^{-4})$	(10^{-4})	$(4.6 * 10^{-4})$

Note: The table presents parameter estimates for respective models based on the full sample. The sample size used for es
stocks and 2000 for SP500 index.
Standard errors, calculated numerically, are given in parentheses.

Table 3: Forecasts evaluation based on MSE loss (full sample)

1-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	5.9626	$0.0940^{\star}$	6.7355	$0.2530^{\star}$	2.5227	0.0400
RT-GARCH-L	5.8989	$0.0990^{\star}$	6.6591	$0.9170^{\star}$	2.0289	$0.6220^{\star}$
RT-GARCH-LF	6.0861	$0.0820^{\star}$	6.6274	$1^{\star}$	1.9370	$1^{\star}$
A-PARCH(2,2)-$St.t$ distr.	5.8152	$0.6330^{\star}$	6.6632	$0.9170^{\star}$	2.6657	0.0020
GARCH(1,1)-$N(0,1)$	5.9074	$0.0990^{\star}$	6.8069	0.0070	2.3780	0.0450
GARCH(1,2)-$N(0,1)$	5.9074	$0.0990^{\star}$	6.8069	0.0070	2.3780	0.0450
GARCH(1,1)-$St.t$ distr.	5.7074	$0.7470^{\star}$	6.8199	0.0030	2.3725	0.0450
GARCH(1,2)-$St.t$ distr.	5.6923	$1^{\star}$	6.9611	0.0030	2.5405	0.0310
Simple NoVaS	7.9097	0	8.6489	0	2.6415	0.0060
Exponential NoVaS	7.9305	0	8.7922	0	2.7714	0.0010

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

5-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	5.4655	$0.0660^{\star}$	6.1762	$0.1130^{\star}$	1.7490	$0.1820^{\star}$
RT-GARCH-L	5.7604	0.0010	6.1410	$0.1130^{\star}$	1.6724	$1^{\star}$
RT-GARCH-LF	7.4039	0.0005	6.8305	0.0005	2.7361	$0.0600^{\star}$
A-PARCH(2,2)-$St.t$ distr.	5.5588	0.0200	6.2182	$0.1130^{\star}$	2.0496	$0.1230^{\star}$
GARCH(1,1)-$N(0,1)$	5.5436	0.0200	6.2221	0.0440	1.7265	$0.1820^{\star}$
GARCH(1,2)-$N(0,1)$	5.5436	0.0020	6.2221	0.0440	1.7265	$0.1820^{\star}$
GARCH(1,1)-$St.t$ distr.	5.2380	$0.1460^{\star}$	6.2346	0.0220	2.1047	$0.1090^{\star}$
GARCH(1,2)-$St.t$ distr.	5.1158	$1^{\star}$	5.9126	$1^{\star}$	2.1059	$0.1090^{\star}$
Simple NoVaS	7.9295	0	8.8844	0	2.8555	0.0220
Exponential NoVaS	7.9810	0	8.9874	0	2.9460	0.0050

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 4: Forecasts evaluation based on MSE loss (full sample)

10-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	5.5838	0.6440*	6.5120	0.5320*	1.8604	1*
RT-GARCH-L	6.7789	0.0020	6.5751	0.0570*	2.3635	0.0020
RT-GARCH-LF	7.1880	0	8.1607	0	2.4339	0.0220
A-PARCH(2,2)-$St.t$ distr.	5.2557	1*	6.3874	0.8660*	2.1500	0.1010*
GARCH(1,1)-$N(0,1)$	5.9061	0.0410	6.7387	0.0110	2.1701	0.0380
GARCH(1,2)-$N(0,1)$	5.9061	0.0410	6.7387	0.0110	2.1701	0.0380
GARCH(1,1)-$St.t$ distr.	5.7090	0.0930*	6.7123	0.0150	2.2999	0.0130
GARCH(1,2)-$St.t$ distr.	5.7043	0.0930*	6.3749	1*	2.4080	0.0020
Simple NoVaS	7.4780	0	9.0686	0	3.0637	0
Exponential NoVaS	7.4882	0	9.1016	0	3.0515	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

15-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	6.2789	0.1290*	6.8308	0.2120*	2.0271	1*
RT-GARCH-L	7.3068	0.0030	7.1004	0.0030	2.8156	0.0070
RT-GARCH-LF	7.7454	0.0010	9.5520	0.0030	2.8458	0.0010
A-PARCH(2,2)-$St.t$ distr.	5.6200	1*	6.5759	1*	2.3697	0.0740*
GARCH(1,1)-$N(0,1)$	6.4643	0.0130	7.4280	0.0020	2.5460	0.0220
GARCH(1,2)-$N(0,1)$	6.4643	0.0130	7.4280	0.0020	2.5460	0.0220
GARCH(1,1)-$St.t$ distr.	7.0099	0.0020	7.3615	0.0020	2.5960	0.0240
GARCH(1,2)-$St.t$ distr.	5.7186	0.7550*	6.8867	0.0290	2.7077	0.0090
Simple NoVaS	8.0218	0	9.1868	0	3.2885	0
Exponential NoVaS	8.0104	0	9.2016	0	3.1058	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 5: Forecasts evaluation based on QLIKE loss (full sample)

1-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.4628	$0.5500^\star$	1.4505	$0.0610^\star$	0.9426	$0.3740^\star$
RT-GARCH-L	1.4531	$1^\star$	1.4315	$1^\star$	0.9471	$0.1550^\star$
RT-GARCH-LF	1.4828	$0.3800^\star$	1.4384	$0.6540^\star$	1.0077	0.0010
A-PARCH(2,2)-$St.t$ distr.	1.5487	0.0120	1.4733	0.0320	0.9322	$1^\star$
GARCH(1,1)-$N(0,1)$	1.5106	0.0320	1.4767	0.0270	0.9488	0.0450
GARCH(1,2)-$N(0,1)$	1.5106	0.0320	1.4767	0.0270	0.9488	0.0450
GARCH(1,1)-$St.t$ distr.	1.5252	0.0400	1.4786	0.0180	0.9494	0.0440
GARCH(1,2)-$St.t$ distr.	1.5277	0.0400	1.4786	0.0180	0.9376	$0.4540^\star$
Simple NoVaS	3.9372	0	3.3669	0	1.4625	0.0005
Exponential NoVaS	3.9361	0	3.3633	0	1.5563	0.0005

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^\star_{95\%}$.

5-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.3935	$1^\star$	1.3760	$1^\star$	0.8957	$0.2950^\star$
RT-GARCH-L	1.4834	0.0140	1.3932	$0.0560^\star$	1.0169	0.0090
RT-GARCH-LF	1.6347	0.0040	1.5198	0.0010	1.2279	0.0040
A-PARCH(2,2)-$St.t$ distr.	1.4492	0.0170	1.4034	$0.0560^\star$	0.8955	$1^\star$
GARCH(1,1)-$N(0,1)$	1.4157	0.0280	1.4163	0.0110	0.9159	0.0310
GARCH(1,2)-$N(0,1)$	1.4157	0.0280	1.4163	0.0110	0.9159	0.0310
GARCH(1,1)-$St.t$ distr.	1.4131	0.0300	1.3948	0.0560	0.9470	0.0120
GARCH(1,2)-$St.t$ distr.	1.4144	0.0300	1.3987	0.0560	0.9425	0.0120
Simple NoVaS	4.1445	0	3.6022	0	1.6679	0.0020
Exponential NoVaS	4.0942	0	3.5691	0	1.8015	0.0020

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^\star_{95\%}$.

Table 6: Forecasts evaluation based on QLIKE loss (full sample)

10-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.4462	$0.3600^{\star}$	1.3985	$1^{\star}$	0.9106	$0.6920^{\star}$
RT-GARCH-L	1.5924	0.0220	1.4476	0.0520	1.1241	0.0090
RT-GARCH-LF	1.8417	0.0130	1.6762	0.0010	1.4544	0.0010
A-PARCH(2,2)-$St.t$ distr.	1.4177	$1^{\star}$	1.4050	$0.8220^{\star}$	0.9081	$1^{\star}$
GARCH(1,1)-$N(0,1)$	1.4864	0.0490	1.4736	0.0010	0.9734	0.0200
GARCH(1,2)-$N(0,1)$	1.4864	0.0490	1.4736	0.0010	0.9734	0.0200
GARCH(1,1)-$St.t$ distr.	1.5101	0.0330	1.4684	0.0010	1.0058	0.0200
GARCH(1,2)-$St.t$ distr.	1.5088	0.0330	1.4014	0.0010	0.9620	0.0380
Simple NoVaS	3.2675	0	3.8264	0	1.9397	0.0005
Exponential NoVaS	3.2776	0	3.7663	0	2.0835	0.0005

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

15-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.4995	$0.1890^{\star}$	1.4267	$0.2660^{\star}$	0.9396	$0.5410^{\star}$
RT-GARCH-L	1.6823	0.0180	1.5039	0.0010	1.1974	0.0080
RT-GARCH-LF	1.7036	0.0090	1.8094	0.0005	1.3169	0.0010
A-PARCH(2,2)-$St.t$ distr.	1.4279	$1^{\star}$	1.4122	$1^{\star}$	0.9352	$1^{\star}$
GARCH(1,1)-$N(0,1)$	1.5533	0.0420	1.5425	0.0050	1.0357	0.0350
GARCH(1,2)-$N(0,1)$	1.5533	0.0420	1.5425	0.0050	1.0357	0.0350
GARCH(1,1)-$St.t$ distr.	1.6097	0.0270	1.5354	0.0020	1.0610	0.0110
GARCH(1,2)-$St.t$ distr.	1.5063	0.0270	1.4534	0.0410	1.0129	$0.0700^{\star}$
Simple NoVaS	3.5635	0	3.8891	0	2.1970	0
Exponential NoVaS	3.5632	0	3.8002	0	2.3442	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 7: Forecasts evaluation based on MSE loss (pre-crisis period)

1-step ahead volatility forecasts

Model	IBM MSE	IBM p_{MCS}	GE MSE	GE p_{MCS}	S&P 500 MSE	S&P 500 p_{MCS}
RT-GARCH	2.2499	$0.8880^{\star}$	2.9478	$0.5310^{\star}$	1.1053	$0.3540^{\star}$
RT-GARCH-L	2.2334	$0.8880^{\star}$	2.9211	$0.5310^{\star}$	1.1198	$0.1810^{\star}$
RT-GARCH-LF	2.3112	$0.3370^{\star}$	2.9234	$0.5310^{\star}$	0.8193	$1^{\star}$
A-PARCH(2,2)-$St.t$ distr.	2.1908	$1^{\star}$	2.8280	$1^{\star}$	0.8508	$0.7260^{\star}$
GARCH(1,1)-$N(0,1)$	2.3644	$0.3000^{\star}$	3.0881	$0.0850^{\star}$	1.1257	0.0470
GARCH(1,2)-$N(0,1)$	2.3644	$0.3000^{\star}$	3.0881	$0.0850^{\star}$	1.1358	0.0470
GARCH(1,1)-$St.t$ distr.	2.2873	$0.7700^{\star}$	3.1117	$0.0690^{\star}$	1.1677	0.0423
GARCH(1,2)-$St.t$ distr.	2.2930	$0.7700^{\star}$	3.1117	$0.0690^{\star}$	1.1716	0.0410
Simple NoVaS	3.5073	0.0010	4.8216	0.0010	1.4206	0.0100
Exponential NoVaS	3.5755	0.0010	4.4485	0.0010	1.4428	0.0100

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

5-step ahead volatility forecasts

Model	IBM MSE	IBM p_{MCS}	GE MSE	GE p_{MCS}	S&P 500 MSE	S&P 500 p_{MCS}
RT-GARCH	1.9340	$0.7160^{\star}$	2.4904	$0.5893^{\star}$	1.0603	$1^{\star}$
RT-GARCH-L	1.6637	$1^{\star}$	2.5724	$0.1220^{\star}$	1.2708	0.0450
RT-GARCH-LF	2.1221	$0.1461^{\star}$	2.8155	0.0430	1.1718	$0.1150^{\star}$
A-PARCH(2,2)-$St.t$ distr.	2.0441	$0.3700^{\star}$	2.2552	$1^{\star}$	1.1618	$0.1150^{\star}$
GARCH(1,1)-$N(0,1)$	2.0740	$0.1460^{\star}$	2.8630	0.0290	1.2890	0.0410
GARCH(1,2)-$N(0,1)$	2.0740	$0.1460^{\star}$	2.8630	0.0290	1.2909	0.0410
GARCH(1,1)-$St.t$ distr.	2.0349	$0.2400^{\star}$	2.8896	0.0220	1.1869	$0.1950^{\star}$
GARCH(1,2)-$St.t$ distr.	2.0311	$0.2421^{\star}$	2.8896	0.0220	1.1300	$0.1950^{\star}$
Simple NoVaS	3.7046	0	4.7204	0.0080	1.5833	0.0150
Exponential NoVaS	3.6552	0	4.8017	0.0090	1.5126	0.0150

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 8: Forecasts evaluation based on MSE loss (pre-crisis period)

10-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	2.0836	0.1470⋆	2.3709	0.0739⋆	1.2375	0.6940⋆
RT-GARCH-L	1.9729	1⋆	2.2761	1⋆	1.1472	1⋆
RT-GARCH-LF	2.3842	0.1000⋆	2.4838	0.0219	1.4586	0.0415
A-PARCH(2,2)-$St.t$ distr.	2.2134	0.1370⋆	2.4967	0.0219	1.3357	0.4150⋆
GARCH(1,1)-$N(0,1)$	2.0265	0.3590⋆	2.7944	0.0180	1.3771	0.0381
GARCH(1,2)-$N(0,1)$	2.0265	0.3590⋆	2.7044	0.0180	1.3788	0.0255
GARCH(1,1)-$St.t$ distr.	2.0823	0.1470⋆	2.8173	0.0150	1.3468	0.2550⋆
GARCH(1,2)-$St.t$ distr.	2.0797	0.1470⋆	2.8173	0.0150	1.3479	0.2410⋆
Simple NoVaS	3.9311	0	5.0234	0	1.6486	0.0380
Exponential NoVaS	3.9594	0	5.0399	0	1.4911	0.0380

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

15-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	2.0443	1⋆	2.4923	0.7170⋆	1.3236	0.7651⋆
RT-GARCH-L	2.3620	0.0130	2.4617	1⋆	1.3178	1⋆
RT-GARCH-LF	2.1540	0.0510⋆	2.7285	0.1210⋆	1.4642	0.0342
A-PARCH(2,2)-$St.t$ distr.	2.1910	0.0460	2.7849	0.0650⋆	1.3413	0.3860⋆
GARCH(1,1)-$N(0,1)$	2.2150	0.0160	2.8142	0.0550⋆	1.4328	0.0120
GARCH(1,2)-$N(0,1)$	2.2150	0.0160	2.8142	0.0550⋆	1.4842	0.0120
GARCH(1,1)-$St.t$ distr.	2.1409	0.1090⋆	2.8321	0.0507⋆	1.3931	0.3420⋆
GARCH(1,2)-$St.t$ distr.	2.1398	0.1220⋆	2.8321	0.0507⋆	1.3994	0.3420⋆
Simple NoVaS	4.2298	0.0060	5.1980	0.0050	1.6308	0.0040
Exponential NoVaS	4.1846	0.0060	5.0943	0.0130	1.4760	0.0040

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 9: Forecasts evaluation based on QLIKE loss (pre-crisis period)

1-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.6258	0.0250	1.6284	0.1046⋆	1.2023	0.0580⋆
RT-GARCH-L	1.5958	0.9310⋆	1.6277	0.2380⋆	1.1762	0.0610⋆
RT-GARCH-LF	1.6602	0.0403	1.6272	0.4090⋆	1.1398	1⋆
A-PARCH(2,2)-$St.t$ distr.	1.5946	1⋆	1.6074	1⋆	1.1509	0.0750⋆
GARCH(1,1)-$N(0,1)$	1.6114	0.2401⋆	1.6591	0.0126	1.2110	0.0180
GARCH(1,2)-$N(0,1)$	1.6114	0.2401⋆	1.6591	0.0126	1.2043	0.0480
GARCH(1,1)-$St.t$ distr.	1.6064	0.2870⋆	1.6643	0.0140	1.1925	0.0350
GARCH(1,2)-$St.t$ distr.	1.6051	0.2870⋆	1.6643	0.0140	1.2084	0.0410
Simple NoVaS	2.6016	0.0077	2.6602	0	1.7735	0.0050
Exponential NoVaS	2.5376	0.0077	2.6812	0	1.7158	0.0010

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

5-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.5482	0.7720⋆	1.5588	0.5490⋆	1.2017	1⋆
RT-GARCH-L	1.5433	1⋆	1.5722	0.0910⋆	1.2427	0.1080⋆
RT-GARCH-LF	1.6091	0.0559⋆	1.5973	0.0420	1.2035	0.5640⋆
A-PARCH(2,2)-$St.t$ distr.	1.5782	0.0853⋆	1.5468	1⋆	1.2340	0.2820⋆
GARCH(1,1)-$N(0,1)$	1.5558	0.4340⋆	1.6172	0.0160	1.2687	0.0408
GARCH(1,2)-$N(0,1)$	1.5558	0.4340⋆	1.6172	0.0160	1.2687	0.0408
GARCH(1,1)-$St.t$ distr.	1.5660	0.1550⋆	1.6121	0.0220	1.2559	0.0440
GARCH(1,2)-$St.t$ distr.	1.5656	0.1550⋆	1.6121	0.0220	1.2559	0.0440
Simple NoVaS	2.7961	0	3.0279	0	1.9897	0.0130
Exponential NoVaS	2.6950	0	2.9584	0.0030	1.8741	0.0160

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 10: Forecasts evaluation based on QLIKE loss (pre-crisis period)

10-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.5591	$0.0920^{\star}$	1.5399	$1^{\star}$	1.2256	$1^{\star}$
RT-GARCH-L	1.5499	$1^{\star}$	1.5493	$0.2980^{\star}$	1.2311	$0.5440^{\star}$
RT-GARCH-LF	1.5598	$0.0520^{\star}$	1.5802	$0.0650^{\star}$	1.2767	0.0453
A-PARCH(2,2)-$St.t$ distr.	1.5990	0.0213	1.6078	0.0060	1.2396	$0.1238^{\star}$
GARCH(1,1)-$N(0,1)$	1.5630	0.0260	1.5917	0.0390	1.3013	0.0238
GARCH(1,2)-$N(0,1)$	1.5630	0.0260	1.5917	0.0390	1.2985	0.0161
GARCH(1,1)-$St.t$ distr.	1.5694	0.0201	1.5957	0.0100	1.2716	$0.0482^{\star}$
GARCH(1,2)-$St.t$ distr.	1.5691	0.0201	1.5957	0.0100	1.2766	$0.0490^{\star}$
Simple NoVaS	3.0539	0	3.3247	0	2.0528	0.0130
Exponential NoVaS	2.8865	0	3.1516	0	1.9589	0.0260

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

15-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.5572	$0.2800^{\star}$	1.5547	$0.3380^{\star}$	1.2486	$1^{\star}$
RT-GARCH-L	1.5473	$0.7015^{\star}$	1.5474	$1^{\star}$	1.2558	$0.4810^{\star}$
RT-GARCH-LF	1.5339	$1^{\star}$	1.6121	0.0290	1.3107	0.0449
A-PARCH(2,2)-$St.t$ distr.	1.6039	0.0352	1.5942	0.0390	1.2982	$0.1565^{\star}$
GARCH(1,1)-$N(0,1)$	1.5672	0.0390	1.5852	0.0450	1.3129	0.0419
GARCH(1,2)-$N(0,1)$	1.5672	0.0390	1.5852	0.0450	1.3105	0.0449
GARCH(1,1)-$St.t$ distr.	1.5695	0.0352	1.5884	0.0420	1.3107	0.0440
GARCH(1,2)-$St.t$ distr.	1.5693	0.0352	1.5884	0.0420	1.3107	0.0440
Simple NoVaS	3.3053	0	3.4334	0.0020	2.0618	0.0040
Exponential NoVaS	3.1063	0.0030	3.1453	0.0010	2.0030	0.0031

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 11: Forecasts evaluation based on MSE loss (crisis and post-crisis period)

1-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	4.5716	0.0020	11.4981	0.3600⋆	1.8311	0.0350
RT-GARCH-L	4.4319	0.6390⋆	11.3606	1⋆	1.6040	0.4080⋆
RT-GARCH-LF	4.4203	1⋆	11.6326	0.0020	1.3981	1⋆
A-PARCH(2,2)-$St.t$ distr.	4.4613	0.3390⋆	11.5223	0.0415	2.3010	0
GARCH(1,1)-$N(0,1)$	4.4750	0.0300	11.6015	0.0050	1.7184	0.0470
GARCH(1,2)-$N(0,1)$	4.4750	0.0300	11.6015	0.0050	1.7184	0.0470
GARCH(1,1)-$St.t$ distr.	4.4878	0.0250	11.5367	0.0400	1.9842	0.0020
GARCH(1,2)-$St.t$ distr.	4.6584	0.0020	11.6527	0.0010	2.4677	0
Simple NoVaS	5.5788	0	12.0917	0	1.9697	0.0020
Exponential NoVaS	5.6237	0	12.1720	0	1.9331	0.0080

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

5-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
RT-GARCH	4.1624	0.0960⋆	10.4028	0.2690⋆	0.8056	1⋆
RT-GARCH-L	4.1081	0.7230⋆	10.2550	1⋆	0.9355	0.0810⋆
RT-GARCH-LF	4.2980	0.0960⋆	11.2363	0	1.6040	0.0010
A-PARCH(2,2)-$St.t$ distr.	3.9632	1⋆	10.6841	0.0110	1.3555	0.0020
GARCH(1,1)-$N(0,1)$	4.2318	0.0310	10.7131	0.0110	1.0701	0.0110
GARCH(1,2)-$N(0,1)$	4.2318	0.0310	10.7131	0.0110	1.0701	0.0110
GARCH(1,1)-$St.t$ distr.	4.2136	0.0400	10.5373	0.0410	1.4371	0.0030
GARCH(1,2)-$St.t$ distr.	4.2136	0.0400	10.5325	0.0410	1.3019	0.0030
Simple NoVaS	5.7350	0	13.3463	0	1.8017	0
Exponential NoVaS	5.7397	0	13.3697	0	1.7878	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 12: Forecasts evaluation based on MSE loss (crisis and post-crisis period)

10-step ahead volatility forecasts

Model	**IBM** MSE	**IBM** p_{MCS}	**GE** MSE	**GE** p_{MCS}	**S&P 500** MSE	**S&P 500** p_{MCS}
RT-GARCH	4.6090	0.0660⋆	10.9616	1⋆	1.1098	1⋆
RT-GARCH-L	4.7146	0.0610⋆	11.1144	0.5020⋆	1.3585	0.1250⋆
RT-GARCH-LF	4.9129	0.0270	11.2357	0.5020⋆	2.9225	0.0005
A-PARCH(2,2)-$St.t$ distr.	4.4016	1⋆	11.1299	0.5020⋆	1.7746	0.0150
GARCH(1,1)-$N(0,1)$	4.6784	0.0420	11.3812	0.0030	1.5144	0.0250
GARCH(1,2)-$N(0,1)$	4.6784	0.0420	11.3812	0.0030	1.5144	0.0250
GARCH(1,1)-$St.t$ distr.	4.6889	0.0420	11.3346	0.0050	2.5789	0.0005
GARCH(1,2)-$St.t$ distr.	4.6519	0.0420	11.3889	0.0030	2.7429	0.0005
Simple NoVaS	3.9311	0	13.5452	0	2.9198	0.0005
Exponential NoVaS	3.9594	0	13.5259	0	2.8922	0.0005

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

15-step ahead volatility forecasts

Model	**IBM** MSE	**IBM** p_{MCS}	**GE** MSE	**GE** p_{MCS}	**S&P 500** MSE	**S&P 500** p_{MCS}
RT-GARCH	4.7423	0.6190⋆	11.3676	1⋆	1.7102	1⋆
RT-GARCH-L	5.0036	0.2210⋆	11.8478	0.0130	2.1316	0.1100⋆
RT-GARCH-LF	5.7171	0.0200	11.4883	0.5870⋆	3.6060	0.0190
A-PARCH(2,2)-$St.t$ distr.	4.6119	1⋆	11.4266	0.5870⋆	2.8358	0.0190
GARCH(1,1)-$N(0,1)$	5.7232	0.0200	11.6886	0.0350	2.1965	0.0410
GARCH(1,2)-$N(0,1)$	5.7232	0.0200	11.6886	0.0350	2.1965	0.0410
GARCH(1,1)-$St.t$ distr.	5.6944	0.0200	11.8160	0.0150	4.3403	0
GARCH(1,2)-$St.t$ distr.	5.7463	0.0200	11.7441	0.0150	4.0890	0
Simple NoVaS	5.9670	0.0100	13.6172	0	3.3435	0.0190
Exponential NoVaS	5.9310	0.0100	13.6538	0	3.3017	0.0190

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 13: Forecasts evaluation based on QLIKE loss (crisis and post-crisis period)

1-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.4640	0.0400	1.4299	0.0680⋆	0.6870	0.2500⋆
RT-GARCH-L	1.4069	0.5150⋆	1.3717	1⋆	0.7022	0.0340
RT-GARCH-LF	1.4054	1⋆	1.4097	0.3960⋆	0.7582	0.0340
A-PARCH(2,2)-$St.t$ distr.	1.4403	0.0640⋆	1.4245	0.0680⋆	0.6751	1⋆
GARCH(1,1)-$N(0,1)$	1.4533	0.0420	1.4573	0.0470	0.7815	0.0150
GARCH(1,2)-$N(0,1)$	1.4533	0.0420	1.4573	0.0470	0.7815	0.0150
GARCH(1,1)-$St.t$ distr.	1.4577	0.0420	1.4410	0.0470	0.6912	0.0340
GARCH(1,2)-$St.t$ distr.	1.5020	0.0380	1.4783	0.0090	0.6980	0.0340
Simple NoVaS	3.4775	0	6.0813	0	1.2339	0
Exponential NoVaS	3.5190	0	6.1866	0	1.3122	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

5-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.3452	0.8750⋆	1.2648	1⋆	0.5482	1⋆
RT-GARCH-L	1.3743	0.0110	1.2444	0.3790⋆	0.6770	0.0200
RT-GARCH-LF	1.4143	0	1.3115	0.0020	0.8759	0.0010
A-PARCH(2,2)-$St.t$ distr.	1.3205	1⋆	1.2974	0.0020	0.5514	0.7820⋆
GARCH(1,1)-$N(0,1)$	1.3635	0.0110	1.2792	0.0370	0.6334	0.0340
GARCH(1,2)-$N(0,1)$	1.3635	0.0110	1.2792	0.0370	0.6334	0.0340
GARCH(1,1)-$St.t$ distr.	1.3552	0.0160	1.2861	0.0020	0.7137	0.0010
GARCH(1,2)-$St.t$ distr.	1.3552	0.0160	1.2794	0.0370	0.5869	0.0440
Simple NoVaS	3.9533	0	6.4805	0	1.3514	0
Exponential NoVaS	3.8880	0	6.6783	0	1.4072	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^{\star}_{95\%}$.

Table 14: Forecasts evaluation based on QLIKE loss (crisis and post-crisis period)

10-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.4416	$0.0580^\star$	1.2917	$1^\star$	0.5835	$1^\star$
RT-GARCH-L	1.4579	0.0070	1.3534	0.0010	0.7893	0.0030
RT-GARCH-LF	1.5612	0	1.3280	$0.0870^\star$	1.0837	0
A-PARCH(2,2)-$St.t$ distr.	1.3864	$1^\star$	1.3162	$0.0870^\star$	0.5922	$0.5530^\star$
GARCH(1,1)-$N(0,1)$	1.4577	0.0050	1.3311	0.0030	0.7282	0.0030
GARCH(1,2)-$N(0,1)$	1.4577	0.0050	1.3311	0.0030	0.7282	0.0030
GARCH(1,1)-$St.t$ distr.	1.4336	$0.0630^\star$	1.3483	0.0020	0.9085	0.0010
GARCH(1,2)-$St.t$ distr.	1.4399	$0.0530^\star$	1.3232	0.0030	0.7230	0.0030
Simple NoVaS	3.7833	0	7.5707	0	1.5315	0
Exponential NoVaS	3.8311	0	6.8522	0	1.5865	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^\star_{95\%}$.

15-step ahead volatility forecasts

	IBM		**GE**		**S&P 500**	
Model	QLIKE	p_{MCS}	QLIKE	p_{MCS}	QLIKE	p_{MCS}
RT-GARCH	1.4751	$0.1410^\star$	1.3163	$1^\star$	0.7297	$1^\star$
RT-GARCH-L	1.4933	$0.1410^\star$	1.4231	0.0030	0.9511	0.0080
RT-GARCH-LF	1.6641	0.0160	1.3511	$0.0810^\star$	1.2957	0.0080
A-PARCH(2,2)-$St.t$ distr.	1.4296	$1^\star$	1.3311	$0.2290^\star$	0.7496	$0.1790^\star$
GARCH(1,1)-$N(0,1)$	1.5646	0.0200	1.3736	0.0160	0.8839	0.0080
GARCH(1,2)-$N(0,1)$	1.5646	0.0200	1.3736	0.0160	0.8839	0.0080
GARCH(1,1)-$St.t$ distr.	1.4537	$0.2020^\star$	1.3980	0.0040	1.1265	0.0080
GARCH(1,2)-$St.t$ distr.	1.4616	$0.1410^\star$	1.3613	0.0300	1.1065	0.0080
Simple NoVaS	3.3053	0	6.7079	0	1.8328	0
Exponential NoVaS	3.1063	0.0030	6.9641	0	1.8260	0

Note: p_{MCS} are the p-values from Model Confidence Set test of Hansen et al.(2011). The p-values that are marked with a $\star$ are those in the model confidence set $\widehat{\mathcal{M}}^\star_{95\%}$.

Table 15: Evaluation of 1-step ahead VaR(5%) forecasts (full sample).

	IBM		**GE**		**S&P 500**	
Model	VR	LR_{cc}	VR	LR_{cc}	VR	LR_{cc}
RT-GARCH	1.08	5.3119	0.8000	5.1842	0.6700	5.6921
RT-GARCH-L	0.9800	5.5283	0.6400	5.4509	0.5800	5.7488
RT-GARCH-LF	0.3600	5.8141	0.4600	5.4396	0.3800	5.7207
A-PARCH(2,2)-$St.t$ distr.	0.7800	5.4278	0.2800	5.7620	0.2000	5.8207
GARCH(1,1)-$N(0,1)$	0.7800	5.4278	0.2200	5.8057	0.2000	5.8207
GARCH(1,2)-$N(0,1)$	0.7600	5.4995	0.3800	5.6248	0.3400	5.7202
GARCH(1,1)-$St.t$ distr.	0.3600	5.8141	0.3000	5.7478	0.1400	5.8668
GARCH(1,2)-$St.t$ distr.	0.5400	5.6659	0.3200	5.7336	0.1800	5.8390

Note: VR denotes the violatio ratio=(# of returns that exceed the VaR(5%) forecast)/(# of the expected violations); LR_{cc} stands for Likelihood Ratio test for conditional coverage, see Christoffersen (1998). Moreover, $LR_{cc} \sim \chi^2_{(2)}$ with critical values 5.99 ($p = 0.05$) and 9.21 ($p = 0.01$).

Appendix.

To derive the eq.(4) observe that λ_t^2 cn be written as follows:

$$\lambda_t^2 = b_{t-1} + \varphi\epsilon_t^2 = b_{t-1} + \varphi\frac{r_t^2}{\lambda_t^2}.$$

Provided that $\lambda_t^2 > 0$, it follows that:

$$\lambda_t^2 = \frac{1}{2}b_{t-1} + \frac{1}{2}\sqrt{b_{t-1}^2 + 4\varphi r_t^2} = \frac{1}{2}b_{t-1} + \frac{1}{2}b_{t-1}\sqrt{1 + \frac{4\varphi r_t^2}{b_{t-1}^2}} \approx \frac{1}{2}b_{t-1} + \frac{1}{2}b_{t-1}\left(1 + \frac{2\varphi r_t^2}{b_{t-1}^2}\right) =$$
$$= b_{t-1} + \frac{\varphi r_t^2}{b_{t-1}} = \frac{\varphi r_t^2}{b_{t-1}} + \alpha + \gamma r_{t-1}^2 + \beta\lambda_{t-1}^2 = \frac{\varphi r_t^2}{b_{t-1}} + \alpha + \gamma r_{t-1}^2 + \beta\left[\frac{\varphi r_{t-1}^2}{b_{t-2}} + \alpha + \gamma r_{t-2}^2 + \beta\lambda_{t-2}^2\right] = \cdots =$$
$$= \frac{\alpha}{1-\beta} + \frac{\varphi r_t^2}{b_{t-1}} + \sum_{j=1}^{\infty}\left(\frac{\beta^j\varphi}{b_{t-1-j}} + \gamma\beta^{j-1}\right) r_{t-j}^2,$$

where in the first line of the derivations we used that for $x << 0$ it holds that $(1+x)^\alpha \approx 1 + \alpha x$.

Proof of Theorem 1. Consider the general model:

$$r_t = \lambda_t\epsilon_t$$

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi\epsilon_t^2,$$

where $\{\epsilon_t\}$ is i.i.d. random variables such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = 1$ with the density f_ϵ. In order to compute $P(r_t \leq c)$ note that the first equation can be rewritten as

$$r_t = \sqrt{\alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi\epsilon_t^2}\epsilon_t$$

such that

$$P(r_t \leq c) = P(\sqrt{\alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi\epsilon_t^2}\epsilon_t \leq c).$$

Since the scaling factor of ϵ_t is positive there is one unique value of d such that

$$\sqrt{\alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi e^2}e \leq c$$

for all $e \leq d$. To obtain d we first square the above equation such that

$$(\alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi d^2)d^2 = c^2 \Leftrightarrow (\alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2)d^2 + \varphi d^4 = c^2. \tag{19}$$

Eq. (19) is a quartic equation in d whenever $\varphi \neq 0$ and is quadratic equation in d whenever $\varphi = 0$ (which is simply the usual GARCH(1,1) case). For quartic equation the solutions are given by:

$$d_{1,2} = \pm\sqrt{\frac{\sqrt{b_{t-1}^2 + 4c^2\varphi} - b_{t-1}}{2\varphi}}$$

$$d_{3,4} = \pm\sqrt{-\frac{\sqrt{b_{t-1}^2 + 4c^2\varphi} + b_{t-1}}{2\varphi}}$$

with $b_{t-1} = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2$. We disregard $x_{3,4}$ since we are only interested in the real valued solutions, such that we have:

$$d(c) = sign(c)\sqrt{\frac{\sqrt{b_{t-1}^2 + 4c^2\varphi} - b_{t-1}}{2\varphi}} \tag{20}$$

and

$$P(r_t \leq c) = \int_{-\infty}^{d(c)} f_\epsilon(x)dx.$$

In order to emphasize the dependence of $d(c)$ on the past information as well as the parameter vector $\theta = (\alpha, \beta, \gamma, \varphi)'$ we adopt the following notation:

$$d(c, b_{t-1}, \theta) = sign(c)\sqrt{\frac{\sqrt{b_{t-1}^2 + 4c^2\varphi} - b_{t-1}}{2\varphi}} \tag{21}$$

The solution to the quadratic equation in d for the case $\varphi = 0$ is given by:

$$d(c, b_{t-1}, \theta) = c/\sqrt{b_{t-1}}, \tag{22}$$

which corresponds to the standard GARCH(1,1) model (as $b_{t-1} = \sigma_t^2$ whenever $\varphi = 0$), for which the conditional density of the returns is just $f_r(r|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{b_{t-1}}} f_\epsilon(\epsilon) = f_\epsilon(\epsilon)/\sigma_t$, where $\mathcal{F}_{t-1}$ denotes the information set up to time $t-1$. To obtain the density in the case $\varphi \neq 0$ we use Leibniz integral rule with variable limits to get:

$$f_r(r|\mathcal{F}_{t-1}) = \frac{\partial P(r_t \le c)}{\partial c} = \frac{\partial d(c, b_{t-1}, \theta)}{\partial c} |_{c=r} \; f_\epsilon(d(r, b_{t-1}, \theta)) =$$

$$\left\{ \frac{\partial sign(c)}{\partial c} \sqrt{\frac{\sqrt{b_{t-1}^2 + 4c^2\varphi} - b_{t-1}}{2\varphi}} + sign(c)\frac{1}{2}\left(\frac{\sqrt{b_{t-1}^2 + 4c^2\varphi} - b_{t-1}}{2\varphi}\right)^{-\frac{1}{2}} \frac{1}{2\varphi} \times \right.$$

$$\left. \times \frac{1}{2}(b_{t-1}^2 + 4c^2\varphi)^{-\frac{1}{2}} 8c\varphi \right\}_{c=r} f_\epsilon(d(r, b_{t-1}, \theta)) =$$

$$= sign(r) r \sqrt{\frac{2\varphi}{(b_{t-1}^2 + 4r^2\varphi)(\sqrt{b_{t-1}^2 + 4r^2\varphi} - b_{t-1})}} f_\epsilon(d(r, b_{t-1}, \theta)) =$$

$$= \frac{|r|}{d(r, b_{t-1}, \theta)\sqrt{b_{t-1}^2 + 4r^2\varphi}} f_\epsilon(d(r, b_{t-1}, \theta)).$$

Remark: Note that $\frac{\partial sign(c)}{\partial c} = 2\delta(c)$, where $\delta(\cdot)$ is a Dirac delta function which is zero everywhere except at 0, where $\delta(0) = \infty$, therefore the above formula holds for $r \neq 0$. Before we calculate the limit of the above equation at $r = 0$, note that $d(r, b_{t-1}, \theta)$ in the denominator iinvolves $sign(r)$, while the numerator involves $|r| = r sign(r)$, we thus can write the density as:

$$f_r(r|\mathcal{F}_{t-1}) = \frac{r}{d(r, b_{t-1}, \theta)\sqrt{b_{t-1}^2 + 4r^2\varphi}} f_\epsilon(d(r, b_{t-1}, \theta))$$

with $d(r, b_{t-1}, \theta) = \sqrt{\frac{\sqrt{b_{t-1}^2 + 4r^2\varphi} - b_{t-1}}{2\varphi}}$. Note that $\epsilon_t = d(r_t; b_{t-1}, \theta_0)$. We now calculate the limit of the density function at $r = 0$. First observe that

$$\lim_{r \to 0} \frac{r}{d(r, b_{t-1}, \theta)} = \frac{r}{\sqrt{\frac{(b_{t-1}^2 + 4\varphi r^2)^{\frac{1}{2}} - b_{t-1}}{2\varphi}}} = \frac{r}{\sqrt{\left(\frac{b_{t-1}^2}{4\varphi^2} + \frac{4\varphi r^2}{4\varphi^2}\right)^{1/2} - \frac{b_{t-1}}{2\varphi}}} =$$

$$= \frac{r}{\sqrt{\frac{b_{t-1}}{2\varphi}\left(\left(1 + \frac{4\varphi r^2}{b_{t-1}^2}\right)^{1/2} - 1\right)}} = \frac{r}{\sqrt{\frac{b_{t-1}}{2\varphi}\left(1 + \frac{1}{2}\frac{4\varphi r^2}{b_{t-1}^2} - 1\right)}} = \sqrt{b_{t-1}}.$$

And as a result we have the following limit

$$\lim_{r \to 0} f_r(r|\mathcal{F}_{t-1}) = \lim_{r \to 0} \frac{r}{d(r, b_{t-1}, \theta)\sqrt{b_{t-1}^2 + 4r^2\varphi}} f_\epsilon(d(r, b_{t-1}, \theta)) = \frac{1}{\sqrt{b_{t-1}}} f_\epsilon(0).$$

The corresponding cumulative distribution function is given by

$$F(r|\mathcal{F}_{t-1}) = \int_{-\infty}^{d(r, b_{t-1}, \theta)} f_\epsilon(x)dx = F_\epsilon\left(d(r, b_{t-1}, \theta)\right).$$

The j^{th} conditional moment of returns can be derived as follows (for the ease of exposition we write $d(r)$ instead of $d(r, b_{t-1}, \theta)$):

$$E\left[r^j|F_{t-1}\right] = \int_{-\infty}^{\infty} r^j f_r(r|\mathcal{F}_{t-1})dr = \int_{-\infty}^{\infty} r^j \frac{r}{d(r)\sqrt{b_{t-1}^2 + 4r^2\varphi}} f_\epsilon(d(r))dr =$$

$$= \int_{-\infty}^{\infty} r^j \frac{\sqrt{b_{t-1} + \varphi d(r)^2}}{\sqrt{b_{t-1}^2 + 4r^2\varphi}} f_\epsilon(d(r))dr = \int_{-\infty}^{\infty} r^j \frac{\sqrt{b_{t-1} + \varphi d(r)^2}}{b_{t-1} + 2\varphi d(r)^2} f_\epsilon(d(r))dr. \quad (23)$$

Now observe that

$$dr = d(d(r))\frac{b_{t-1} + 2\varphi d(r)^2}{\sqrt{b_{t-1} + \varphi d(r)^2}}.$$

Thus with a change of variable of integration eq.(23) can be written as

$$E\left[r^j|F_{t-1}\right] = \int_{-\infty}^{\infty} r^j \frac{\sqrt{b_{t-1}+\varphi d(r)^2}}{b_{t-1}+2\varphi d(r)^2} f_\epsilon(d(r))dr = \int_{-\infty}^{\infty} r^j f_\epsilon(d(r))d(d(r)) =$$
$$= \int_{-\infty}^{\infty} d(r)^j \left(b_{t-1}+\varphi d(r)^2\right)^{j/2} f_\epsilon(d(r))d(d(r)) = \int_{-\infty}^{\infty} d(r)^j \left(b_{t-1}(1+\frac{\varphi d(r)^2}{b_{t-1}})\right)^{j/2} f_\epsilon(d(r))d(d(r)) =$$
$$= \int_{-\infty}^{\infty} d(r)^j b_{t-1}^{j/2}\left((1+\frac{\varphi d(r)^2}{b_{t-1}})\right)^{j/2} f_\epsilon(d(r))d(d(r)) = \int_{-\infty}^{\infty} d(r)^j b_{t-1}^{j/2}\left((1+\frac{j}{2}\frac{\varphi d(r)^2}{b_{t-1}})\right) f_\epsilon(d(r))d(d(r)) =$$
$$= b_{t-1}^{j/2}\left[\int_{-\infty}^{\infty} d(r)^j f_\epsilon(d(r))d(d(r)) + \frac{j\varphi}{2b_{t-1}}\int_{-\infty}^{\infty} d(r)^{j+2} f_\epsilon(d(r))d(d(r))\right] =$$
$$= b_{t-1}^{j/2}\left[E\left[d(r)^j\right] + \frac{j\varphi}{2b_{t-1}} E\left[d(r)^{j+2}\right]\right],$$

where in the third line of this derivations we used the first-order Taylor approximation and the fact that $d(r)$ is symmetric around zero, $E[d(r)] = 0$. ■

Proof of Theorem 2. The general model is given by:

$$r_t = \lambda_t \epsilon_t \tag{24}$$

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma r_{t-1}^2 + \varphi\epsilon_t^2. \tag{25}$$

Since the error term ϵ_t is i.i.d, it is then obvious that the error process $(\epsilon_t)_{t\in\mathbb{Z}}$ is always strictly stationary and ergodic. Thus, $(r_t)_{t\in\mathbb{Z}}$ is a strictly stationary process if and only if $(\lambda_t)_{t\in\mathbb{Z}}$ is strictly stationary. Therefore, the task of deriving the strict stationarity conditions for the whole process $(r_t, \lambda_t)_{t\in\mathbb{Z}}$ can be reduced to deriving strict stationarity conditions for $(\lambda_t^2)_{t\in\mathbb{Z}}$, given by eq.(25).

Let's now express $(\lambda_t^2)_{t\in\mathbb{Z}}$ solely in terms of the error process $(\epsilon_t)_{t\in\mathbb{Z}}$. Repeatedly substituting for λ_{t-1}^2 in eq.(25) , we have:

$$\lambda_t^2 = \lambda_0 \prod_{i=1}^{t}(\beta+\gamma\epsilon_{t-i}^2) + \sum_{i=0}^{t}\left(\prod_{j=0}^{i-1}(\beta+\gamma\epsilon_{t-j-1}^2)\right)(\alpha+\varphi\epsilon_{t-i}), \qquad t \geq 2. \tag{26}$$

In order for eq.(26) to be well defined we need either to assume the trivial σ-algebra $\mathcal{F}_0$ (and a probability measure μ_0) for the starting value λ_0^2 or to assume that the system extends infinitely far into the past. We proceed by implementing the former approach, defining:

$$\mathbb{P}[\lambda_0^2 \in \Gamma] = \mu_0(\Gamma) \qquad \forall \Gamma \in \mathcal{B} \quad \text{and} \quad \mu_0\left((0,\infty)\right) = 1, \tag{27}$$

where $\mathcal{B}$ denotes the Borel sets on $[0,\infty)$. In order to find strict stationarity conditions of λ_t^2 we next rewrite eq.(25) in the form of the stochastic difference equation $Y_{t+1} = A_t Y_t + B_t$, where Y_t, A_t and B_t are given by:

$$A_t = \beta + \gamma \epsilon_t^2, \qquad B_t = \alpha + \varphi \epsilon_{t+1}^2 \qquad \text{and} \qquad Y_t = \lambda_t^2 \tag{28}$$

Since sequences $(A_t)_{t\in\mathbb{N}}$ and $(B_t)_{t\in\mathbb{N}}$ are measurable transformations of the strictly stationary and ergodic process $(\epsilon_t)_{t\in\mathbb{N}}$ we can make use of the Theorem 3.5.8 of Stout (1974) to claim that these sequences are strictly stationary and ergodic as well as the sequence $\Psi = (A_t, B_t)_{t\in\mathbb{N}}$. If we rewrite eq.(26) in terms of eq.(28), it follows that $(Y_t)_{t\in\mathbb{N}} = (\lambda_t^2)_{t\in\mathbb{N}}$ is the solution of the stochastic difference equation $Y_{t+1} = A_t Y_t + B_t$. Every such solution then should satisfy the following representation:

$$Y_{t+1} = A_t Y_t + B_t = A_t A_{t-1} Y_{t-1} + A_t B_{t-1} + B_t = A_t A_{t-1} A_{t-2} Y_{t-2} + A_t A_{t-1} B_{t-2} + A_t B_{t-1} + B_t =$$
$$= \cdots = \left(\prod_{i=0}^{t} A_{t-i}\right) Y_0 + \sum_{i=0}^{t} \left(\prod_{j=0}^{i-1} A_{t-j}\right) B_{t-i}, \tag{29}$$

with the usual convention that $\prod_{j=0}^{-1} A_{t-j} = 1$ for the product over an empty index set. Let's denote by Y an arbitrary $\mathbb{R}$-valued random variable, which is defined on the same probability space as Ψ. Note that Y and Ψ should not necessarily be independent. The solution $y_t(Y,\Psi)$ of eq.(29) is then given by:

$$y_t(Y,\Psi) = \left(\prod_{i=0}^{t-1} A_i\right) Y_0 + \sum_{i=0}^{t-1}\left(\prod_{j=t-i}^{t-1} A_j\right) B_{t-i-1}.$$

We have shown earlier that the sequence $\Psi = (A_t, B_t)$ is strictly stationary and ergodic, we can now apply Theorem 1 of Brandt (1986) to deduce that $y_t(\Psi) = \sum_{i=0}^{\infty}\left(\prod_{j=n-i}^{t-1} A_j\right) B_{t-i-1}$, $t \in \mathbb{N}$ is strictly stationary solution if and only if the following conditions are satisfied:

$$\mathbb{P}(A_0 = 0) > 0$$

or

$$-\infty \leqq E \log|A_0| < 0 \qquad\qquad E\,(\log|B_0|)^+ < \infty,$$

where $x^+ = \max(0,x)$ for $x \in \mathbb{R}$. Plugging in the expressions for A_0 and B_0, given by eq.(28) we get the following strict stationarity conditions:

$$-\infty \leqq E \log\left|\beta + \gamma\epsilon_0^2\right| < 0 \qquad\qquad E\left(\log\left|\alpha + \varphi\epsilon_0^2\right|\right)^+ < \infty,$$

in addition to requiring that $\beta > 0, \gamma > 0$ and $\varphi \neq 0$. ■

Proof of Theorem 3.

The result follows directly from combining eq. (6) and eq.(7). ■

Proof of Theorem 4.

The proof follows directly from Theorem 3 and the fact that $E[r_t^2] = E[\lambda_t^2] + \varphi(E[\epsilon_t^4] - 1)$. For proving the last claim of the Theorem 4 notice that since $E[r_t|\mathcal{F}_{t-1}] = 0$ for all t, then $cov(r_t, r_s|\mathcal{F}_{t-1}) = E\left[r_t r_s|\mathcal{F}_{t-1}\right]$ for $s < t$, which we will calculate by direct integration against the density of r_t, i.e.

$$E\left[r_t r_s|\mathcal{F}_{t-1}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_t r_s f_{r_t r_s}(r_t,r_s)dr_t dr_s=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r_t r_s f_{r_t}(r_t)f_{r_s}(r_s)dr_t dr_s=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d(r_t)\left(b_{t-1}+\varphi d(r_t)^2\right)^{1/2} f_\epsilon(d(r_t))d(r_s)\left(b_{s-1}+\varphi d(r_s)^2\right)^{1/2} f_\epsilon(d(r_s))d(d(r_t))d(d(r_s))=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d(r_t)\left(b_{t-1}(1+\frac{\varphi d(r_t)^2}{b_{t-1}})\right)^{1/2} f_\epsilon(d(r_t))d(r_s)\left(b_{s-1}(1+\frac{\varphi d(r_s)^2}{b_{s-1}})\right)^{1/2} f_\epsilon(d(r_s))d(d(r_t)d(d(r_s))=$$

$$=\int_{-\infty}^{\infty} d(r_t)b_{t-1}^{1/2}\left((1+\frac{\varphi d(r_t)^2}{b_{t-1}})\right)^{1/2} f_\epsilon(d(r_t))d(d(r_t))\int_{-\infty}^{\infty} d(r_s)b_{s-1}^{1/2}\left((1+\frac{\varphi d(r_s)^2}{b_{s-1}})\right)^{1/2} f_\epsilon(d(r_s))d(d(r_s))=$$

$$=\int_{-\infty}^{\infty} d(r_t)b_{t-1}^{1/2}\left((1+\frac{1}{2}\frac{\varphi d(r_t)^2}{b_{t-1}})\right) f_\epsilon(d(r_t))d(d(r_t))\int_{-\infty}^{\infty} d(r_s)b_{s-1}^{1/2}\left((1+\frac{1}{2}\frac{\varphi d(r_s)^2}{b_{s-1}})\right) f_\epsilon(d(r_s))d(d(r_s))=$$

$$=b_{t-1}^{1/2}\left[\int_{-\infty}^{\infty} d(r_t)f_\epsilon(d(r_t))d(d(r_t))+\frac{j\varphi}{2b_{t-1}}\int_{-\infty}^{\infty} d(r_t)^3 f_\epsilon(d(r_t))d(d(r_t))\right]\times$$

$$b_{s-1}^{1/2}\left[\int_{-\infty}^{\infty} d(r_s)f_\epsilon(d(r_s))d(d(r_s))+\frac{j\varphi}{2b_{s-1}}\int_{-\infty}^{\infty} d(r_s)^3 f_\epsilon(d(r_s))d(d(r_s))\right]=$$

$$=b_{t-1}^{1/2}b_{s-1}^{1/2}\left[E\left[d(r_t)\right]+\frac{\varphi}{2b_{t-1}}E\left[d(r_t)^3\right]\right]\left[E\left[d(r_s)\right]+\frac{\varphi}{2b_{t-1}}E\left[d(r_s)^3\right]\right]$$

Given that the error term $d(r)$ is symmetric around zero, i.e. its mean and skewness are zero, we get exactly zero in the last equation above. ■

Proof of Theorem 5.

$$E[r_t^4]=E[\lambda_t^4 r_t^4]=(\alpha+\beta\lambda_{t-1}^2+\gamma r_{t-1}^2+\varphi\epsilon_t^2)(\alpha+\beta\lambda_{t-1}^2+\gamma r_{t-1}^2+\varphi\epsilon_t^2)\epsilon_t^4=\alpha^2E[\epsilon_t^4]+2\alpha\beta E[\lambda_{t-1}^2]E[\epsilon_t^4]+$$
$$+2\alpha\gamma E[r_{t-1}^2]E[\epsilon_t^4]+2\alpha\varphi E[\epsilon_t^6]+\beta^2E[\lambda_{t-1}^4]E[\epsilon_t^4]+2\beta\gamma E[\lambda_{t-1}^2]E[r_{t-1}^2]E[\epsilon_t^4]+$$
$$+2\beta\varphi E[\lambda_{t-1}^2]E[\epsilon_t^6]+\gamma^2E[r_{t-1}^4]E[\epsilon_t^4]+2\varphi\gamma E[r_{t-1}^2]E[\epsilon_t^6]+\varphi^2E[\epsilon_t^8].$$

Substituting eq.(7) in the equation above and rearranging we get:

$$(\alpha+4\alpha\gamma\varphi)E[\epsilon_t^4]+(2\alpha\beta+2\alpha\gamma+4\beta\gamma\varphi)E[\lambda_{t-1}^2]E[\epsilon_t^4]+(2\alpha\varphi+4\varphi^2\gamma)E[\epsilon_t^6]+$$
$$+(\beta^2+2\beta\gamma)E[\lambda_{t-1}^4]E[\epsilon_t^4]+(2\beta\varphi+2\gamma\varphi)E[\lambda_{t-1}^2]E[\epsilon_t^6]+\gamma^2E[r_t^4]E[\epsilon_t^4]+\varphi^2E[\epsilon_t^8]$$

If r_t is fourth order stationary ($E[r_t^4] = E[r_{t-1}^4]$), then

$$E[r_1^4] = \Bigg[(\alpha + 4\alpha\gamma\varphi)E[\epsilon_t^4] + (2\alpha\beta + 2\alpha\gamma + 4\beta\gamma\varphi)E[\lambda_{t-1}^2]E[\epsilon_t^4] + (2\alpha\varphi + 4\varphi^2\gamma)E[\epsilon_t^6] + \\ + (\beta^2 + 2\beta\gamma)E[\lambda_{t-1}^4]E[\epsilon_t^4] + (2\beta\varphi + 2\gamma\varphi)E[\lambda_{t-1}^2]E[\epsilon_t^6] + \varphi^2 E[\epsilon_t^8]\Bigg]\Bigg[1 - \gamma^2 E[r_t^4]E[\epsilon_t^4]\Bigg]^{-1}$$

Since $E[r^4]$ must be positive, γ^2 must also satisfy (in addition to case (1) or case (2)):

$$1 - \gamma^2 E[\epsilon_t^4] > 0 \quad \Leftrightarrow \quad \gamma^2 < \frac{1}{E[\epsilon_t^4]}. \qquad \blacksquare$$

Proof of Theorem 6.

We start by writing down the RT-GARCH model with leverage and feedback:

$$r_t = \lambda_t \varepsilon_t$$

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \gamma_1 r_{t-1}^2 \mathbb{1}_{(r_t>0)} + \gamma_2 r_{t-1}^2 \mathbb{1}_{(r_t\leq 0)} + \varphi_1 \epsilon_t^2 \mathbb{1}_{(\epsilon_t>0)} + \varphi_2 \epsilon_t^2 \mathbb{1}_{(\epsilon_t\leq 0)}. \tag{30}$$

Denoting by $\kappa := E\left[\varepsilon_t^4\right]$ and $\eta := \kappa - 1$ and following the same steps as in the proof of Theorems 3 and 4 (see details in in the Supplementary Material to this paper) we have:

$$E[r_t^2] = \alpha + \beta E[\lambda_{t-1}^2] + \gamma_1 E[r_{t-1}^2 | r_t > 0] + \gamma_2 E[r_{t-1}^2 | r_t \leq 0] + \varphi_1 E\left[\epsilon_t^4 | \epsilon_t > 0\right] + \varphi_2 E\left[\epsilon_t^4 | \epsilon_t \leq 0\right] \tag{31}$$

and

$$E[\lambda_t^2] = \alpha + \beta E[\lambda_{t-1}^2] + \gamma_1 E[r_{t-1}^2 | r_t > 0] + \gamma_2 E[r_{t-1}^2 | r_t \leq 0] + \varphi_1 E\left[\epsilon_t^2 | \epsilon_t > 0\right] + \varphi_2 E\left[\epsilon_t^2 | \epsilon_t \leq 0\right]. \tag{32}$$

Combining eq.(31)-(32) then yields:

$$E[r_t^2] = E[\lambda_t^2] + (\varphi_1 + \varphi_2)\Bigg(E\left[\varepsilon_t^4\right] - 1\Bigg) \tag{33}$$

In addition we also have that the unconditional first moment of λ_t^2 is

$$E[\lambda_1^2] = \frac{\alpha + (\varphi_1 + \varphi_2)\left[\eta(\gamma_1 + \gamma_2) + 1\right]}{1 - (\beta + \gamma_1 + \gamma_2)}. \tag{34}$$

Using eq.(31) and (32) we can write:

$$E\left[\lambda_{t+1}^2 | \mathcal{F}_t\right] = \alpha + (\varphi_1 + \varphi_2)\left[\eta(\gamma_1 + \gamma_2) + 1\right] + (\beta + \gamma_1 + \gamma_2) E\left[\lambda_t^2 | \mathcal{F}_t\right]. \tag{35}$$

Now note that from eq.(34) we have:

$$\alpha + (\varphi_1 + \varphi_2)\left[\eta(\gamma_1 + \gamma_2) + 1\right] = \left[1 - (\beta + \gamma_1 + \gamma_2)\right] E\left[\lambda_1^2\right],$$

which, when substituted back into eq.(35), together with eq.(33) provides us with the formula in Theorem 6. ■