ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis in Quantitative Finance

July 4, 2021

A Critical Review and Enhanced Version of Real-Time GARCH

Stijn van Zoggel

484049

Supervisor: dr. Rutger-Jan Lange

Second assessor: dr. Onno Kleen

Abstract

This paper corrects and improves on the Real-Time GARCH (RT-GARCH) volatility model by Smetanina (2017a), which combines the advantages of both GARCH and stochastic volatility (SV) models in a unified framework. We namely find that her contributory out-of-sample volatility forecasts are spurious due to incorrect proofs. We correct these proofs and introduce an enhanced version of RT-GARCH, which better captures asymmetric effects of returns on volatility. Furthermore, we compare the corrected and enhanced models to SV models, as Smetanina (2017a) refrains from this. After constructing 1-, 5-, 15- and 30-days ahead volatility forecasts for two stock indices, we find that the corrected RT-GARCH forecasts are not convincingly better than those of conventional GARCH models, but better than SV models. Our enhanced model on the other hand, yields superior forecasts compared to all other models.

1 Introduction

Volatility is a widely used proxy for the risk of financial assets and plays a crucial role within many areas of finance, from risk management to asset allocation and derivative pricing. Over the last decades, much research has been conducted on reliable estimation and forecasting models for the volatility of asset returns. The largest challenge faced in this line of research is the correct representation of the non-Gaussian and clustering behavior returns exhibit. Non-Gaussian, because extreme asset returns occur more often than expected under normality, resulting in a more fat-tailed and peaked distribution (Mandelbrot, 1963). In the same paper it is noted that periods of large absolute returns alternate with periods of small absolute returns, suggesting volatility is not constant.

Two well-established and competing models have been developed to capture these fat tails and volatility clustering properties of asset returns, namely Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models and stochastic volatility (SV) models. Smetanina (2017a) introduces a Real-Time GARCH (RT-GARCH) model that combines the advantages of both GARCH and SV models in a unified framework. Her model obtains better volatility forecasts than GARCH models, however we find that these results are partly spurious due to wrong proofs and mathematical notation. In this paper we will therefore address her mistakes and propose the Enhanced RT-GARCH (ERT-GARCH) model. Moreover, we also compare RT-GARCH and ERT-GARCH to the more complex SV models, as Smetanina (2017a) refrains from this.

GARCH models (Engle, 1982; Bollerslev, 1986; Nelson, 1991; Ding et al., 1993; Zakoian, 1994, among others) assume that volatility is a function entirely determined by past information, making it one-step ahead deterministic. This observation-driven nature makes GARCH models popular among researchers, because parameter estimates are easily obtained using estimation methods. Especially quasi-maximum likelihood estimation (QMLE) is consistent and asymptotically normal under mild conditions (Francq & Zakoian, 2007). Nevertheless, there is also a downside to the complete predetermination of volatility by GARCH models, as Politis (2007) remarks that exclusion of current information leads to inadequate forecasts. Consequently, GARCH models adapt slowly during turmoil times in financial markets, whereas real volatility changes rapidly.

SV models on the other hand, characterize volatility as a stochastic process. Pioneered by Clark (1973) and Taylor (1986), this is done by not only allowing for past internal information, but also for current external information in the form of unobserved random shocks. These shocks make the models more responsive to abrupt changes in volatility, which is confirmed by Geweke (1994), who finds that SV models fit return data better than GARCH models. Unfortunately, the apparent advantages of SV models over GARCH models come at a price. Likelihood-based estimation of SV models namely requires integration of the latent volatility process, which can only be effectuated using complicated numerical methods (Fridman & Harris, 1998). Various estimation approaches have been established, usually based on the method of moments (Andersen et al., 1997), QMLE (Harvey et al., 2012) or Markov Chain Monte Carlo (Jacquier et al, 1994).

Until recently, all volatility models were either categorized as parameter-driven or observation-driven. In an attempt to find a compromise between the simplicity of GARCH models and the flexibility of SV models, Smetanina (2017a) proposes the Real-Time GARCH model. This new model efficiently uses all available internal information, by modeling volatility as a mixture of past as well as current information. More specifically, a deterministic current return innovation is added to the volatility process, which still allows for the simple QMLE framework that GARCH models are popular for. Smetanina (2017a) shows that RT-GARCH nests the GARCH model as a special case, while volatility is modeled in the spirit of SV models, with perfectly correlated random errors for return and volatility.

Compared to the standard GARCH(1,1) model, enlarging the information set of the volatility process results in a quicker adjustment to the new level of volatility after a sudden jump. Another important advantage of RT-GARCH over GARCH, is that it allows for a time-varying kurtosis of the conditional distribution of returns. This provides a better empirical fit to return data, especially in the tails. All together, Smetanina (2017a) finds that the RT-GARCH model outperforms standard GARCH models in terms of both short- and long-term volatility forecasts.

The aforementioned results obtained by the RT-GARCH model of Smetanina (2017a) are very promising, as the best of both volatility worlds is combined. However, some of the proofs on which RT-GARCH is based are wrong and a crucial misconstruction is made for the RT-GARCH model that distinguishes between negative and positive returns. Besides, it is only investigated whether the RT-GARCH model improves on existing GARCH models, leaving comparison with SV models for further research. The main purpose of this paper will hence be to place the RT-GARCH model in full perspective, by pointing to its flaws, revising its proofs, suggesting an Enhanced RT-GARCH model and comparing it to both GARCH and SV models. Eventually this will add value to the ongoing research on optimal volatility model selection.

RT-GARCH is a relatively new model and to our knowledge only Ding (2020) has conducted follow-up research on the topic at hand. He concludes that RT-GARCH and GARCH weakly converge to the same diffusion process, meaning RT-GARCH performs at least as well as GARCH, but only when the return data is sampled at high enough frequency.

In an empirical study we first estimate the RT-GARCH, ERT-GARCH, GARCH and SV models in a full-sample analysis for two global stock indices. The current return innovation in RT-GARCH and ERT-GARCH models results in a better in-sample fit compared to the conventional GARCH and SV models. Subsequently we estimate all models within a rolling window procedure to create

1-, 5-, 15- and 30- day ahead volatility forecasts. Considering forecast accuracy based on the MSPE loss function, we conclude that RT-GARCH forecasts are not convincingly better than those of conventional GARCH models, but better than SV models. Our enhanced model on the other hand, yields superior forecasts compared to all other models.

The rest of this paper is organized as follows. Section 2 first gives a critical review of the RT-GARCH framework, proposes the ERT-GARCH framework and presents the SV model. In this section we will also elaborate on their estimation and volatility forecasting procedures. Subsequently, Section 3 describes the data set for two stock indices and presents empirical results for the volatility models, including a forecast evaluation. Finally, Section 4 states our conclusion.

2 Methodology

2.1 A Critical Review of RT-GARCH

The RT-GARCH model assumes that the additional "current information" can be characterized with the current return scaled by its volatility. First observed by Black (1976), negative news shocks generally have a larger effect on volatility than positive news shocks. This is called the "leverage" effect. Similarly, negative returns have a larger influence on volatility than positive returns, which is referred to as the "feedback effect". To allow for such leverage and feedback effects, Smetanina (2017a) defines a general RT-GARCH-Leverage-Feedback (RT-GARCH-LF) model, which boils down to the following joint process:

$$\text{RT-GARCH-LF} = \begin{cases} r_t = \lambda_t z_t, & (1) \\ \lambda_t^2 = \sigma_t^2 + \left(\phi_1 \mathbb{1}_{z_t \leq 0} + \phi_2 \mathbb{1}_{z_t > 0}\right) z_t^2, & (2) \\ \sigma_t^2 = \alpha + \beta \lambda_{t-1}^2 + \left(\gamma_1 \mathbb{1}_{r_t \leq 0} + \gamma_2 \mathbb{1}_{r_t > 0}\right) r_{t-1}^2, & (3) \end{cases}$$

with $\alpha > 0$, $(\beta, \gamma_1, \gamma_2, \phi_1, \phi_2) \geq 0$ and where r_t is the (demeaned) return at time t, λ_t^2 the volatility process and z_t are i.i.d. standardized residuals with density function $f_z(\cdot)$, such that $E(z_t) = 0$ and $E(z_t^2) = 1$. Note that it is possible to incorporate current information in this manner, since we do not require knowledge of the unobserved future returns, but only of the first and second moments of the standardized errors. These errors are squared to satisfy the necessary condition $\lambda_t^2 > 0$.

A downside of the RT-GARCH-LF model is that the indicator functions in equation (3) consider r_t instead of r_{t-1}, such that σ_t is not one-step ahead deterministic. Later in this section we elucidate why this has some serious implications for the credibility of the model. Moreover, the conventional asymmetric Threshold-GARCH (T-GARCH) is not nested as a result of the uncommon indicator functions. In Section 2.2 we solve this problem by adjusting the indicator functions $\mathbb{1}_{r_t\leq}$ and $\mathbb{1}_{r_t>0}$ to $\mathbb{1}_{r_{t-1}\leq 0}$ and $\mathbb{1}_{r_{t-1}>0}$. Another drawback of the model by Smetanina (2017a) is that it does not account for the asymmetric distribution of r_t in case leverage or feedback effects are present. Including a constant parameter μ in equation (1) should therefore improve the model specification, which is done is Section 2.2 as well.

From the general model the RT-GARCH-Leverage (RT-GARCH-L) model can be obtained by setting $\phi_1 = \phi_2 = \phi$, while the RT-GARCH model imposes the additional restriction $\gamma_1 = \gamma_2 = \gamma$. Whenever $\phi_1 = \phi_2 = 0$ and $\gamma_1 = \gamma_2 = \gamma$, we get the symmetric GARCH model. The whole RT-GARCH framework can thus be summarized as:

$$\text{RT-GARCH-LF} = \begin{cases} \text{RT-GARCH-L}, & \text{if } \theta = (\alpha,\ \beta,\ \gamma,\ \gamma, \phi_1, \phi_2), \\ \text{RT-GARCH}, & \text{if } \theta = (\alpha,\ \beta,\ \gamma,\ \gamma,\ \phi,\ \phi), \\ \text{T-GARCH}, & \text{not possible}, \\ \text{GARCH}, & \text{if } \theta = (\alpha,\ \beta,\ \gamma,\ \gamma,\ 0,\ 0). \end{cases}$$

Next, we state some important results from Smetanina (2017a). We point out the errors in her proofs and revise them using similar derivations as in Lange (2021). The results hold for the various parameter restrictions as stated above.

Result 1: Let $\left(r_t, \lambda_t^2\right)$ evolve according to the ERT-GARCH-LF process. Then, assuming $\beta + \left(\frac{\gamma_1+\gamma_2}{2}\right) < 1$, the process λ_t^2 is weakly stationary and its first unconditional moment is given by

$$\lambda^2 := E\left[\lambda_t^2\right] = \frac{\alpha + \left(\dfrac{\phi_1+\phi_2}{2}\right) + \left(\dfrac{\phi_1+\phi_2}{2}\right)\left(\dfrac{\gamma_1+\gamma_2}{2}\right)\left(E\left[z_t^4\right]-1\right)}{1-\beta-\left(\dfrac{\gamma_1+\gamma_2}{2}\right)}. \tag{4}$$

The proof of this result is given in Appendix A.1. Note that this expression divides $\gamma_1 + \gamma_2$ and $\phi_1 + \phi_2$ by two, while the unconditional mean $E[\lambda_1^2]$ in Smetanina (2017a, Section 5, Theorem 6) does not. In her proof, which can be found in Smetanina (2017b, Section 2), the expectations of equations (23) and (24) are miscalculated. As shown in our proof, it should namely be that $E\left[\left(\gamma_1 \mathbb{1}_{r_t\leq 0} + \gamma_2 \mathbb{1}_{r_t>0}\right) r_{t-1}^2\right] = \frac{1}{2}(\gamma_1+\gamma_2)E[r_{t-1}^2]$ instead of $(\gamma_1+\gamma_2)E[r_{t-1}^2]$. In the same way, it should be that $E\left[\left(\phi_1 \mathbb{1}_{\epsilon_t\leq 0} + \phi_2 \mathbb{1}_{\epsilon_t>0}\right) \epsilon_{t-1}^2\right] = \frac{1}{2}(\phi_1+\phi_2)$ instead of $(\phi_1+\phi_2)$.

Result 2: Denote with $\mathcal{I}_{t-1}$ the information set up to time t–1 and with $f_z(\cdot)$ the probability density function (pdf) of z_t. The partly problematic conditional pdf of the returns is then given by

$$f_r\left(r|\mathcal{I}_{t-1}\right) = \frac{\sqrt{L(r,\mathcal{I}_{t-1})}}{L(r,\mathcal{I}_{t-1}) + F(r)Z(r,\mathcal{I}_{t-1})^2} \cdot f_z(Z(r,\mathcal{I}_{t-1})), \tag{5}$$

where

$$\begin{aligned} F(r) &= \phi_1 \mathbb{1}_{r_t \leq 0} + \phi_2 \mathbb{1}_{r_t > 0}, \\ Z(r,\mathcal{I}_{t-1}) &= \frac{r}{\sqrt{L(r,\mathcal{I}_{t-1})}}, & r \in \mathbb{R}, \\ L(r,\mathcal{I}_{t-1}) &= \frac{\sigma_t^2 + \sqrt{\sigma_t^4 + 4F(r)r^2}}{2}, & r \in \mathbb{R}, \end{aligned}$$

and σ_t^2 is given by equation (3). For the proof of this result we refer to Sections 1.3 and 1.4 of Lange (2021), where only $r_t - \mu$ should be replaced with r_t. The conditional pdf is problematic for the RT-GARCH-LF model, because σ_t^2 then quietly incorporates r_t and is in fact not known based on the information set $\mathcal{I}_{t-1}$. This discrepancy harms the main idea of only using current information from the shocks and might artificially increase the log-likelihood, the function of which follows almost directly from the conditional pdf.

Result 3: The QML estimator $\hat{\theta}$ of the true parameter $\theta_0 = (\alpha_0, \beta_0, \gamma_{1,0}, \gamma_{2,0}, \phi_{1,0}, \phi_{2,0})$ is given by $\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ L_T(\theta)$, where $L_T(\theta)$ is the log-likelihood function that, assuming a standard Gaussian distribution for z_t, can be written as:

$$L_T(\theta) = \sum_{t=1}^{T} \frac{1}{2}\left[-\log(2\pi) + 2\log\left(\frac{\sqrt{\lambda_t^2}}{\lambda_t^2 + F(r_t)Z(r_t,\mathcal{I}_{t-1})^2}\right) - Z(r_t,\mathcal{I}_{t-1})^2\right], \tag{6}$$

where λ_t^2 is given by equation (2). Smetanina and Wu (2019) show that the QML estimator is consistent and asymptotically normally distributed for the symmetric RT-GARCH model.

Result 4: Let $\left(r_t, \lambda_t^2\right)$ evolve according to the RT-GARCH-LF process. Then the k-step ahead predicted volatility, $k \geq 1$, is given by the following formula:

$$\hat{\sigma}_{t+k|t} = \sqrt{E_t[r_{t+k}^2]} = \sqrt{\lambda^2 + \left(\beta + \frac{\gamma_1+\gamma_2}{2}\right)^{k-1}\left(E_t[\lambda_{t+1}^2] - \lambda^2\right) + \left(\frac{\phi_1+\phi_2}{2}\right)\left(E\left[z_t^4\right]-1\right)}, \tag{7}$$

where $E_t[\lambda_{t+1}^2] = \alpha + \beta\lambda_t^2 + \frac{\gamma_1+\gamma_2}{2} r_t^2 + \frac{\phi_1+\phi_2}{2}$ is unknown at time t and λ^2 is the unconditional mean given in equation (4). The proof of this result is given in Appendix A.2. Once more this equation is different from the square root of the k-step ahead forecast in Smetanina (2017a, Section 5, Theorem 6). Firstly, $\gamma_1 + \gamma_2$ and $\phi_1 + \phi_2$ are not divided by two in her paper, caused by the same miscalculated expectations as mentioned in Result 1. Secondly, we raise the second term to the power $k-1$, while Smetanina (2017a) raises it to the power k. Setting $k = 1$ quickly confirms that her formula is incorrect, since the correct result in Smetanina (2017a, eq. 18) can never be obtained. The problem is that the proof of Smetanina (2017a, Appendix) calculates $E_t[\lambda_{t+k}^2] = \lambda^2 + (\beta+\gamma)E_t[\lambda_{t+k-1}^2 - \lambda^2]$ for $k \geq 1$, but this should be for $k \geq 2$ instead. Our reasoning can be found after equation (27) in Appendix A.2.

2.2 An Enhanced Version of RT-GARCH: ERT-GARCH

As we previously mentioned, the RT-GARCH framework by Smetanina (2017a) has some flaws. First, the distribution of r_t is no longer symmetric for the RT-GARCH-L and RT-GARCH-LF models. To improve the model specification, we include a constant parameter μ in equation (1). Another crucial flaw of the RT-GARCH framework is that σ_t is not one-step ahead deterministic for the RT-GARCH-LF model. To solve this problem, we adjust the indicator functions $\mathbb{1}_{r_t \leq \mu}$ and $\mathbb{1}_{r_t > \mu}$ to $\mathbb{1}_{r_{t-1} \leq \mu}$ and $\mathbb{1}_{r_{t-1} > \mu}$. Taking both adjustments together, we define the new Enhanced RT-GARCH-LF (ERT-GARCH-LF) model as follows:

$$\text{ERT-GARCH-LF} = \begin{cases} r_t = \mu + \lambda_t z_t, & (8) \\ \lambda_t^2 = \sigma_t^2 + \left(\phi_1 \mathbb{1}_{z_t \leq 0} + \phi_2 \mathbb{1}_{z_t > 0}\right) z_t^2, & (9) \\ \sigma_t^2 = \alpha + \beta\lambda_{t-1}^2 + \left(\gamma_1 \mathbb{1}_{r_{t-1} \leq \mu} + \gamma_2 \mathbb{1}_{r_{t-1} > \mu}\right)(r_{t-1} - \mu)^2, & (10) \end{cases}$$

with $\alpha > 0$, $(\beta, \gamma_1, \gamma_2, \phi_1, \phi_2) \geq 0$ and z_t i.i.d. with density function $f_z(\cdot)$, such that $E(z_t) = 0$ and $E(z_t^2) = 1$. When $\phi_1 \neq \phi_2$, Lange (2021) shows that μ is still the median of the returns r_t, but no longer the mean. The following models are nested by this general form:

$$\text{ERT-GARCH-LF} = \begin{cases} \text{ERT-GARCH-L}, & \text{if } \theta = (\mu,\ \alpha,\ \beta,\ \gamma,\ \gamma, \phi_1, \phi_2), \\ \text{ERT-GARCH}, & \text{if } \theta = (\mu,\ \alpha,\ \beta,\ \gamma,\ \gamma,\ \phi,\ \phi), \\ \text{ET-GARCH}, & \text{if } \theta = (\mu,\ \alpha,\ \beta, \gamma_1, \gamma_2,\ 0,\ 0), \\ \text{E-GARCH}, & \text{if } \theta = (\mu,\ \alpha,\ \beta,\ \gamma,\ \gamma,\ 0,\ 0). \end{cases}$$

Next, we state the important results for the ERT-GARCH model framework, which hold for the nested models as well. The proofs can be found in Lange (2021).

Result 5: Let $\left(r_t, \lambda_t^2\right)$ evolve according to the ERT-GARCH-LF process. Then, assuming $\beta + \left(\frac{\gamma_1+\gamma_2}{2}\right) < 1$, the process λ_t^2 is weakly stationary and its first unconditional moment is given by

$$\lambda^2 := E\left[\lambda_t^2\right] = \frac{\alpha + \left(\dfrac{\phi_1+\phi_2}{2}\right) + \dfrac{\gamma_1}{2}\left(\phi_1 E[z_t^4] - \dfrac{\phi_1+\phi_2}{2}\right) + \dfrac{\gamma_2}{2}\left(\phi_2 E[z_t^4] - \dfrac{\phi_1+\phi_2}{2}\right)}{1 - \beta - \left(\dfrac{\gamma_1+\gamma_2}{2}\right)}. \tag{11}$$

Result 6: Denote with $\mathcal{I}_{t-1}$ the information set up to time t–1 and with $f_z(\cdot)$ the pdf of z_t. The non-problematic conditional pdf of the returns is then given by

$$f_r\left(r|\mathcal{I}_{t-1}\right) = \frac{\sqrt{L(r,\mathcal{I}_{t-1})}}{L(r,\mathcal{I}_{t-1}) + F(r)Z(r,\mathcal{I}_{t-1})^2} \cdot f_z(Z(r,\mathcal{I}_{t-1})), \tag{12}$$

where

$$\begin{aligned} F(r) &= \phi_1 \mathbb{1}_{r_t \le \mu} + \phi_2 \mathbb{1}_{r_t > \mu}, & \\ Z(r,\mathcal{I}_{t-1}) &= \frac{r-\mu}{\sqrt{L(r,\mathcal{I}_{t-1})}}, & r \in \mathbb{R}, \\ L(r,\mathcal{I}_{t-1}) &= \frac{\sigma_t^2 + \sqrt{\sigma_t^4 + 4F(r)(r-\mu)^2}}{2}, & r \in \mathbb{R}. \end{aligned}$$

and σ_t^2 now is a known quantity based on the information set $\mathcal{I}_{t-1}$, given by equation (10). For the proof of this result we refer to sections 1.2 and 1.3 of Lange (2021). The next result follows almost directly from this pdf.

Result 7: The QML estimator $\hat{\theta}$ of the true parameter $\theta_0 = (\mu_0, \alpha_0, \beta_0, \gamma_{1,0}, \gamma_{2,0}, \phi_{1,0}, \phi_{2,0})$ is given by $\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ L_T(\theta)$, where $L_T(\theta)$ is the log-likelihood function that, assuming a standard Gaussian distribution for z_t, can be written as:

$$L_T(\theta) = \sum_{t=1}^{T} \frac{1}{2}\left[-\log(2\pi) + 2\log\left(\frac{\sqrt{\lambda_t^2}}{\lambda_t^2 + F(r_t)Z(r_t,\mathcal{I}_{t-1})^2}\right) - Z(r_t,\mathcal{I}_{t-1})^2\right], \tag{13}$$

where λ_t^2 is given by equation (9).

Result 8: Let $\left(r_t, \lambda_t^2\right)$ evolve according to the ERT-GARCH-LF process. Then the k-step ahead predicted volatility, $k \geq 1$, is given by the following formula:

$$\begin{aligned}\hat{\sigma}_{t+k|t} &= \sqrt{E_t[r_{t+k}^2]} \\ &= \sqrt{\mu^2 + \lambda^2 + \left(\beta + \frac{\gamma_1+\gamma_2}{2}\right)^{k-1}\left(\sigma_{t+1}^2 + \left(\frac{\phi_1+\phi_2}{2}\right) - \lambda^2\right) + \left(\frac{\phi_1+\phi_2}{2}\right)\left(E\left[z_t^4\right]-1\right)},\end{aligned} \tag{14}$$

where σ_{t+1}^2 is known at time t according to equation (10) and λ^2 is the unconditional mean given in equation (11).

2.3 A Competing Model: Stochastic Volatility

A simple, non-linear and stationary SV model is defined by the following state space model:

$$\mathrm{SV}_{\mathrm{NL}} = \begin{cases} r_t = \epsilon_t \exp(h_t/2) & (15) \\ h_t = \omega + \delta h_{t-1} + \eta_t, & (16) \end{cases}$$

where $h_t = \ln(\sigma_t^2)$, $\epsilon_t \sim$ i.i.d. $(0,1)$, $\eta_t \sim$ i.i.d. $(0, \sigma_\eta^2)$ and $|\delta| < 1$. The assumption that unconditional volatility is log-normally distributed is supported in empirical studies (see for example Andersen et al., 2001) and ensures that σ_t^2 is always positive. As explained in Smetanina (2017a, Section 1.2), RT-GARCH and $\mathrm{SV}_{\mathrm{NL}}$ models in their simplest form differ from each other in the sense that the processes of RT-GARCH are driven by one common shock z_t, while those of $\mathrm{SV}_{\mathrm{NL}}$ are driven by two independent shocks ϵ_t and η_t.

Equation (16) is the state equation, which describes how volatility evolves from time $t-1$ to time t. Equation (15) is the observation equation, which expresses how the underlying state is transformed into measurable returns. Filtering methods can estimate the parameters of this model from a series of noisy measurements of r_t. The most popular filtering method is the Kalman filter (Kalman, 1960), a recursive algorithm that provides an estimate of h_t given the information known at time $t-1$, driven by the prediction error. However, the Kalman filter requires the state space model to be linear, while the current SV model is highly non-linear. In section 2.3.1 we therefore describe how the SV model can be linearized and estimated via QMLE of the Kalman filter.

Recent work by Lange (2020) presents the Bellman filter, a generalization of the Kalman filter that does not require the SV model to be linear. It differs from the Kalman filter as it is driven by the score of the observation density, making it more robust when the observation noise is heavy tailed. Using a simulation study, Lange (2020) shows that the Bellman filter outperforms the

Kalman filter in terms of the mean absolute error of one-step ahead volatility forecasts. We check if these enhancements can also be seen in an empirical study, by also estimating the SV model via QMLE of the Bellman filter. For the sake of brevity we do not elaborate on the Bellman filtering method in this paper. The algorithm and its technicalities can be found in Lange (2020).

2.3.1 Kalman Filter

Nelson (1988) shows that the non-linear SV_{NL} model can be put into a linear state space form by taking logarithms of the squared returns in equation (15):

$$\ln(r_t^2) = \ln(\epsilon_t^2 \exp(h_t)) = \ln(\epsilon_t^2) + h_t = E\left[\ln(\epsilon_t^2)\right] + h_t + \xi_t,$$

where $\xi_t = \ln(\epsilon_t^2) - E\left[\ln(\epsilon_t^2)\right]$ is a non-Gaussian, white noise process with statistical properties depending on the distribution of ϵ_t. Assuming $\epsilon_t \sim$ i.i.d. $N(0,1)$, derivations from Abramowitz & Stegun (1972) evince that $\ln(\epsilon_t^2) \dot{\sim}$ i.i.d.$(-1.27,\, \pi^2/2)$. Now the SV_{NL} model can be written as:

$$\text{SV}_{\text{L}} = \begin{cases} \ln(r_t^2) = -1.27 + h_t + \xi_t & (17) \\ h_t = \omega + \delta h_{t-1} + \eta_t, & (18) \end{cases}$$

where $\xi_t \dot{\sim}$ i.i.d.$(0,\, \pi^2/2)$. This model is in a linear state space form, such that QMLE of the Kalman filter can be used to obtain parameter estimates. Denoting with $\hat{h}_{t|t} = E[h_t|\mathcal{I}_t]$ our best conditional volatility estimate at time t and with $P_{t|t} = E[(h_t - \hat{h}_{t|t})^2]$ the uncertainty of this estimate, we define the recursive equations of the Kalman filter in Table 1. The filter is initialized with the unconditional moments of h_t.

Table 1: The Kalman filter for the SV_{L} model

Step	Computation
Initialize	$\hat{h}_{0\|0} = \omega(1-\delta)^{-1}$ and $P_{0\|0} = \sigma_\eta^2(1-\delta^2)^{-1}$. Set $t = 1$.
Predict	While $t \leq T$, $\hat{h}_{t\|t-1} = \omega + \delta \hat{h}_{t-1\|t-1}$ $P_{t\|t-1} = \omega + \delta^2 P_{t-1\|t-1} + \sigma_\eta^2$
Update	$\hat{h}_{t\|t} = \hat{h}_{t\|t-1} + P_{t\|t-1}\left(P_{t\|t-1} + \pi^2/2\right)^{-1}\left(\ln(y_t^2) + 1.27 - h_{t\|t-1}\right)$ $P_{t\|t} = P_{t\|t-1} + P_{t\|t-1}^2\left(P_{t\|t-1} + \pi^2/2\right)^{-1}$. Set $t = t + 1$ and return to the "Predict" step.
Smooth	$\hat{h}_{T\|T} = \hat{h}_{T\|T}$ and $P_{T\|T} = P_{T\|T}$. Set $t = t - 1$. While $t \geq 0$, backwardly calculate $\hat{h}_{t\|T} = \hat{h}_{t\|t} + \delta P_{t\|t}(P_{t+1\|t})^{-1}(h_{t+1\|T} - h_{t+1\|t})$ $P_{t\|T} = P_{t\|t} + \delta^2 P_{t\|t}^2 (P_{t+1\|t})^{-2}(P_{t+1\|T} - P_{t+1\|t})$.

Result 9: The QML estimator $\hat{\theta}$ of the true parameter $\theta_0 = (\omega_0, \delta_0, \sigma_{\eta,0})$ is given by $\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}}\ L_T(\theta)$, where $L_T(\theta)$ is the log-likelihood function that, using prediction error decomposition (Harvey, 1981), can be written as:

$$L_T(\theta) = \sum_{t=1}^{T} \frac{1}{2} \left[-\log(2\pi) - \log\left(P_{t|t-1} + \pi^2/2\right) - \left(\ln(y_t^2) + 1.27 - h_{t|t-1}\right)^2 \left(P_{t|t-1} + \pi^2/2\right)^{-1} \right]. \tag{19}$$

Wooldridge (1991) proves that QMLE will yield consistent and asymptotically normal parameter estimates if the SV model correctly specifies the first two conditional moments, which is the case here.

Result 10: Having found the model parameter estimates via either the Kalman or Bellman filter, we can recursively find the expression for the k-step ahead volatility prediction:

$$\begin{aligned} \hat{\sigma}_{t+k|t} &= \sqrt{E_t[r_{t+k}^2]} \\ &= \sqrt{\exp\left(\delta^k h_{t|t} + \omega\left(\frac{1-\delta^k}{1-\delta}\right) + \frac{\sigma_\eta^2}{2}\left(\frac{1-\delta^{2k}}{1-\delta^2}\right)\right)}. \end{aligned} \tag{20}$$

where $h_{t|t}$ is the estimate from the filter at time t. The proof of this result is given in Appendix A.3.

2.4 Forecast Evaluation

In sections 2.1 to 2.3 we defined 11 models: GARCH, E-GARCH, ET-GARCH, RT-GARCH, RT-GARCH-L, RT-GARCH-LF, ERT-GARCH, ERT-GARCH-L, ERT-GARCH-LF, SV-Kalman and SV-Bellman. At first we define the forecasting error at time t as $e_{t+k|t} = RV_{t+k} - \hat{\sigma}_{t+k|t}$, where RV_{t+k} is the realized volatility, a proxy for the unobserved conditional volatility. We then test for the following desirable properties of our volatility forecasts:

- *Unbiasedness:* the forecast errors have zero mean, i.e. H_0: $E[e_{t+k|t}] = 0$. This can be tested using a simple Student's t-test.

- *Efficiency:* the volatility forecasts should correctly explain the realized variance in a Mincer-Zarnowitz regression, i.e. H_0: $\alpha_0 = 0$ and $\alpha_1 = 1$ when estimating $RV_{t+k} = \alpha_0 + \alpha_1 \hat{\sigma}_{t+k|t} + \varepsilon_{t+k}$. This can again be tested using a Student's t-test.

- *Accuracy:* the Mean Squared Prediction Error (MSPE) should be as small as possible and at least smaller than the standard deviation of the realized volatility, i.e. we want $E\left[e_{t+k|t}^2\right] =$

$Var[e_{t+k|t}] + E[e_{t+k|t}]^2 < Var[RV_{t+k}]$. Patton (2011) shows that the MSPE is a robust loss function for which is does not matter if one compares $\hat{\sigma}_{t+k|t}$ with the true conditional volatility σ_{t+k} or some consistent proxy such as the realized volatility RV_{t+k}.

- *Relative accuracy:* as in Smetanina (2017a) we test if the models fall in the Model Confidence Set (MCS) $\mathcal{M}^*$ of Hansen et al. (2011). Using the MSPE as loss function, we get that $\mathcal{M}^*_{1-\alpha} = \{i \in \mathcal{M}_0 : E[e^2_{i,t+k|t} - e^2_{j,t+k|t}] \leq 0, \text{ for all } j \in \mathcal{M}_0\}$ with probability $(1-\alpha)$. Here i and j are indices of models in the initial set $\mathcal{M}_0$. For the exact specification of the MCS algorithm and its test we refer to Hansen et al. (2011).

3 Empirical Application

3.1 Data

We use the daily logarithmic open-to-close returns of the Standard and Poor's 500 (S&P 500) and Euro Stoxx 50 (Stoxx 50) indices from the realized library at Oxford-Man Institute of Quantitative Finance. The sample period is from 2 January 2004 to 31 March 2021, amounting to respectively 4328 and 4402 observations for the S&P 500 and Stoxx 50 returns. Summary statistics of the asset returns are given in Table 2. From this table it can be seen that the returns show non-Gaussian behavior, as the kurtoses are way higher than 3 and the Jarque-Bera test rejects the null hypothesis of normality. To capture volatility clustering and keep a consistent number of observations on which the RT-GARCH, ERT-GARCH and SV parameters are based, we use a rolling window estimation procedure. This differs slightly from the papers by Smetanina (2017a) and Ding (2021), who use an expanding window procedure. Following recommendations by Ng & Lam (2006), we take a window of 1000 observations, which results in volatility forecasts from 1 April 2008 to 31 March 2021.

Table 2: Summary statistics of the daily logarithmic open-to-close returns (in percent) over the period 2 January 2004 - 31 March 2021.

Returns	Sample size	Mean	St.dev.	Skewness	Kurtosis	JB
S&P 500	4328	0.0170	1.1870	-0.3362	14.3458	23292***
Stoxx 50	4402	-0.0066	1.1870	-0.5293	11.8500	14571***

Note: JB represents the outcome of the Jarque-Bera test, $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

For out-of-sample forecast evaluation, a proxy for the conditional volatility is needed. We will take this to be the 5-minute realized variance from intraday high-frequency data, obtained from the Oxford-Man Institute of Quantitative Finance as well.

3.2 Full-Sample Analysis

We estimate the parameters of the 11 volatility models over the full sample period from 2 January 2004 to 31 March 2021. The estimates, Akaike Information Criterion (AIC) and log-likelihood of the RT-GARCH and ERT-GARCH models for the S&P 500 and Stoxx 50 returns are given in Table 3. The ϕ parameters differ significantly from zero for all RT-GARCH and ERT-GARCH models, indicating that inclusion of a current news shock improves the model specification. From the γ and ϕ estimates of the RT-GARCH-L and RT-GARCH-LF, it can be seen that feedback and leverage effects are apparent. This especially holds for the Stoxx 50 returns, with ϕ_1 and γ_1 respectively having a much larger effect on volatility than ϕ_2 and γ_2. If we subsequently look at the ϕ parameters of ERT-GARCH-L and ERT-GARCH-LF, it can be seen that the leverage effect is still apparent, but to a lesser extent than in the asymmetric RT-GARCH models. On the other hand, the asymmetry in γ increases substantially when μ is included, even to the point where the γ_2 estimates are insignificant.

Based on the AIC and log-likelihood, we would select RT-GARCH-LF as the best model for the S&P 500 index. However, recall from Section 2.1 that the log-likelihood of this model might be artificially increased as a result from σ_t that is not one-step ahead deterministic. Estimating the quasi-deterministic version of RT-GARCH, namely ERT-GARCH with μ fixed to zero, leads to an AIC and log-likelihood of respectively 2.3333 and -5043 for S&P 500. These values are larger and smaller than their RT-GARCH-LF equivalents, so we conclude that the in-sample fit of RT-GARCH-LF by Smetanina (2017a) can indeed be unrealistically good. For this reason, we regard RT-GARCH-L and especially ERT-GARCH-LF as better models. The results for Stoxx 50 are less ambiguous: we select ERT-GARCH-LF as the best model.

Generally, the RT-GARCH and ERT-GARCH models have a better model specification than the conventional GARCH models (GARCH, E-GARCH and ET-GARCH), for which the results are in Table 6 in Appendix B. The AIC of SV-Bellman, displayed in Table 7 in Appendix B, is noticeably lower than the AIC's of the conventional GARCH models, but mostly higher than those of the RT-GARCH and ERT-GARCH models. The AIC of SV-Kalman, given in the same table, is by far the largest of all models.

Table 3: Parameter estimates, AIC and log-likelihood of the RT-GARCH and ERT-GARCH models for S&P 500 and Stoxx 50 over the period 2 January 2004 - 31 March 2021. Numerically calculated standard errors are in parentheses.

	μ	α	β	γ_1	γ_2	ϕ_1	ϕ_2	AIC	L_T
S&P 500									
RT-GARCH	-	0.0000	0.8405***	0.1133***	-	0.0266***	-	2.3646	-5113
		(0.0041)	(0.0150)	(0.0125)		(0.0034)			
RT-GARCH-L	-	0.0000	0.8586***	0.0790***	-	0.0743***	0.0018	2.3308	-5039
		(0.0029)	(0.0117)	(0.0094)		(0.0086)	(0.0015)		
RT-GARCH-LF	-	0.0000	0.8628***	0.1398***	0.0359***	0.0673***	0.0028*	2.3233	-5022
		(0.0029)	(0.0111)	(0.0171)	(0.0097)	(0.0081)	(0.0017)		
ERT-GARCH	0.0584***	0.0000	0.8350***	0.1156***	-	0.0281***	-	2.3604	-5103
	(0.0091)	(0.0041)	(0.0155)	(0.0127)		(0.0035)			
ERT-GARCH-L	0.0480***	0.0000	0.8397***	0.1092***	-	0.0348***	0.0215***	2.3586	-5098
	(0.0009)	(0.0041)	(0.0155)	(0.0127)		(0.0046)	(0.0034)		
ERT-GARCH-LF	0.0307***	0.0000	0.8535***	0.1958***	0.0000	0.0265***	0.0235***	2.3311	-5038
	(0.0096)	(0.0035)	(0.0139)	(0.0196)	(0.0132)	(0.0038)	(0.0034)		
Stoxx 50									
RT-GARCH	-	0.0000	0.8618***	0.0841***	-	0.0469***	-	2.7897	-6136
		(0.0078)	(0.0167)	(0.0120)		(0.0062)			
RT-GARCH-L		0.0000	0.8972***	0.0436***	-	0.1198***	0.0000	2.7554	-6060
		(0.0050)	(0.0111)	(0.0079)		(0.0134)	(0.0032)		
RT-GARCH-LF	-	0.0000	0.9008***	0.0853***	0.0112*	0.1052***	0.0020	2.7476	-6042
		(0.0046)	(0.0097)	(0.0126)	(0.0072)	(0.0122)	(0.0029)		
ERT-GARCH	0.0562***	0.0000	0.8526***	0.0895***	-	0.0497***	-	2.7841	-6123
	(0.0120)	(0.0081)	(0.0179)	(0.0128)		(0.0064)			
ERT-GARCH-L	0.0316***	0.0000	0.8674***	0.0695***	-	0.0738***	0.0302***	2.7770	-6106
	(0.0010)	(0.0066)	(0.0176)	(0.0122)		(0.0089)	(0.0051)		
ERT-GARCH-LF	0.0189***	0.0000	0.8641***	0.1731***	0.0000	0.0472***	0.0331***	2.7457	-6036
	(0.0001)	(0.0057)	(0.0144)	(0.0199)	(0.0125)	(0.0073)	(0.0053)		

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Whenever ϕ_2 and/or γ_2 is not estimated, it is actually the case that $\phi_1 = \phi$ and/or $\gamma_1 = \gamma$.

3.3 Out-of-Sample Forecast Evaluation

We now estimate the 11 models using a moving window of 1000 days over the period from 1 April 2008 to 31 March 2021. Every iteration we construct 1-, 5-, 15- and 30-day ahead volatility forecasts according to equations (7), (14) and (20) and we check for their desirable properties as explained in Section 2.4.

Unbiasedness:

Table 4 presents the mean forecast errors of the S&P 500 and Stoxx 50 volatility forecasts. From this table it can be seen that the null hypothesis of unbiased forecasts is rejected with 1% significance for all models except SV-Kalman with $k = 1$ (both stock indices) and $k = 15$ (only Stoxx 50).

We conclude that our models generally overpredict the magnitude of (realized) volatility, as all biased models have a negative mean error. This is in line with conclusions from previous research on biased GARCH forecasts (see for example Ederington & Guan, 2010). Relatively speaking, the ERT-GARCH-LF model consistently obtains a small forecast error.

Table 4: Mean forecast errors of the S&P 500 and Stoxx 50 volatility forecasts over the period 1 April 2008 - 31 March 2021. The standard errors are given in parentheses.

	$k=1$		$k=5$		$k=15$		$k=30$	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
S&P 500								
GARCH	-0.1143***	(0.0070)	-0.1358***	(0.0085)	-0.1662***	(0.0102)	-0.1852***	(0.0113)
E-GARCH	-0.1153***	(0.0070)	-0.1380***	(0.0085)	-0.1702***	(0.0102)	-0.1906***	(0.0113)
ET-GARCH	-0.1150***	(0.0066)	-0.1330***	(0.0085)	-0.1541***	(0.0102)	-0.1604***	(0.0112)
RT-GARCH	-0.1001***	(0.0071)	-0.1134***	(0.0086)	-0.1226***	(0.0103)	-0.1211***	(0.0113)
RT-GARCH-L	-0.1060***	(0.0071)	-0.1251***	(0.0088)	-0.1354***	(0.0106)	-0.1282***	(0.0117)
RT-GARCH-LF	-0.1068***	(0.0070)	-0.1367***	(0.0086)	-0.1663***	(0.0104)	-0.1733***	(0.0116)
ERT-GARCH	-0.1020***	(0.0070)	-0.1166***	(0.0086)	-0.1260***	(0.0103)	-0.1235***	(0.0114)
ERT-GARCH-L	-0.1016***	(0.0072)	-0.1106***	(0.0087)	-0.1179***	(0.0103)	-0.0485***	(0.0114)
ERT-GARCH-LF	-0.0776***	(0.0067)	-0.0948***	(0.0085)	-0.0662***	(0.0101)	-0.0695***	(0.0112)
SV-Kalman	0.0004	(0.0084)	-0.0378***	(0.0095)	-0.0901***	(0.0109)	-0.1267***	(0.0118)
SV-Bellman	-0.0483***	(0.0073)	-0.1241***	(0.0087)	-0.2511***	(0.0105)	-0.3529***	(0.0117)
Stoxx 50								
GARCH	-0.0874***	(0.0076)	-0.0964***	(0.0088)	-0.1119***	(0.0106)	-0.1268***	(0.0115)
E-GARCH	-0.0876***	(0.0076)	-0.0974***	(0.0088)	-0.1143***	(0.0106)	-0.1304***	(0.0115)
ET-GARCH	-0.0725***	(0.0072)	-0.0825***	(0.0088)	-0.0996***	(0.0109)	-0.1145***	(0.0118)
RT-GARCH	-0.0771***	(0.0078)	-0.0845***	(0.0091)	-0.0846***	(0.0108)	-0.0757***	(0.0116)
RT-GARCH-L	-0.0726***	(0.0080)	-0.0917***	(0.0093)	-0.1090***	(0.0110)	-0.1113***	(0.0118)
RT-GARCH-LF	-0.0722***	(0.0076)	-0.0925***	(0.0090)	-0.1138***	(0.0107)	-0.1203***	(0.0117)
ERT-GARCH	-0.0782***	(0.0077)	-0.0862***	(0.0090)	-0.0862***	(0.0108)	-0.0771***	(0.0116)
ERT-GARCH-L	-0.0725***	(0.0079)	-0.0713***	(0.0092)	-0.0932***	(0.0111)	-0.0991***	(0.0123)
ERT-GARCH-LF	-0.0592***	(0.0072)	-0.0654***	(0.0088)	-0.0541***	(0.0118)	-0.0518***	(0.0115)
SV-Kalman	0.0693	(0.0090)	0.0399***	(0.0098)	-0.0082	(0.0109)	-0.0422***	(0.0116)
SV-Bellman	-0.0047***	(0.0080)	-0.0728***	(0.0092)	-0.1879***	(0.0108)	-0.2792***	(0.0118)

Note: Reject H_0: $E[e_{t+k|t}] = 0$ with $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Efficiency:

Table 8 in Appendix C presents the results of the Mincer-Zarnowitz regressions RV_{t+k} on $\hat{\sigma}_{t+k|t}$, for the 1-, 5-, 15- and 30-day forecast horizons. Since either one or both of the estimates for α_0 and α_1 differs significantly from zero or one, we note that none of the models is efficient, for none of the horizons. The determination coefficients R^2 for the 1-day ahead forecasts are substantial however, with the volatility forecasts explaining up to 71.8% (S&P 500) and 63.8% (Stoxx 50) of the variance in the realized volatility. Especially ET-GARCH and ERT-GARCH-LF score well based on this metric. The R^2 decreases gradually as the forecast horizon increases, with values still up to 0.537

and 0.470 for $k = 5$, but values up to 0.194 and 0.094 at $k = 30$. Also, the more the horizon increases, the more the conventional GARCH models dominate RT-GARCH, ERT-GARCH and SV models in explaining realized volatility.

(Relative) Accuracy:

The MSPE's of the volatility forecasts for the S&P 500 and Stoxx 50 indices are depicted in Table 5, together with the p-values from $\hat{\mathcal{M}}_{0.95}$, the MCS that contains the best model with 95% confidence. Only the short-term forecasts (1- and 5-day ahead) are accurate, because their MSPE's are substantially lower than 0.4921 and 0.4632, respectively the variances of the S&P 500 and Stoxx 50 realized volatility. Overall, we conclude that our Enhanced RT-GARCH model significantly outperforms the original RT-GARCH models by Smetanina (2017a), the conventional GARCH models and SV models when it comes to accurate forecasts. This can be inferred from our two main findings given below.

Firstly, we find that none of the RT-GARCH models convincingly outperforms the conventional GARCH models, which contradicts the main conclusion in Smetanina (2017a). This contradiction with Smetanina (2017a) is most likely caused by her wrong expressions for the unconditional mean and volatility forecasts as explained in Section 2.1. For Stoxx 50, the RT-GARCH models namely never outperform GARCH, except for RT-GARCH-LF at $k = 1$. For S&P 500, the MSPE of RT-GARCH is always lower than for GARCH, but still never lower than for ET-GARCH up to $k = 30$. RT-GARCH-LF only performs better than GARCH for $k = 1$, meaning the artificially improved fit of RT-GARCH-LF in the full-sample analysis of Section 3.2 does not lead to superior accuracy in the out-of-sample volatility forecasts.

Secondly, we find that ERT-GARCH or ERT-GARCH-LF is the most accurate model, depending on the stock index and forecast horizon. ERT-GARCH-LF is the most accurate model for the S&P 500 returns, as this model always lies in the MCS and obtains the lowest MSPE for all forecast horizons. For the Stoxx 50 returns, the ERT-GARCH model always yields the lowest MSPE, but is left out of the MCS at 1- and 30-day ahead horizons. In these cases ERT-GARCH-LF, the model we would expect to do well based on the full-sample analysis of Section 3.2, is the better option. Fully resorting to ERT-GARCH-LF is not a good idea however, since its 15-day ahead MSPE is ranked eleventh, with a value that is even higher than the MSPE of the 30-day ahead forecasts. We do not have a sufficient explanation for this striking observation, such that ERT-GARCH-LF cannot be pointed out as the single most accurate model.

At the other extreme, we conclude that SV-Kalman produces the least accurate short-term forecasts, while SV-Bellman takes over this position for long-term horizons (15- and 30-day ahead). The poor accuracy is as expected, because the SV models have the least number of parameters and do not incorporate leverage or feedback effects.

Table 5: MSPE of the S&P 500 and Stoxx 50 volatility forecasts over the period 1 April 2008 - 31 March 2021.

	k=1		k=5		k=15		k=30	
	MSPE	p_{MCS}	MSPE	p_{MCS}	MSPE	p_{MCS}	MSPE	p_{MCS}
S&P 500								
GARCH	0.1746 (8)	0.0007	0.2564 (5)	0.2259⋆	0.3659 (6)	0.0609⋆	0.4490 (6)	1.0000⋆
E-GARCH	0.1731 (5)	0.0007	0.2565 (6)	0.3299⋆	0.3673 (7)	1.0000⋆	0.4514 (7)	0.0000
ET-GARCH	<u>0.1544</u> (1)	0.0536⋆	0.2523 (2)	0.0881⋆	0.3600 (2)	0.0583⋆	0.4379 (5)	0.0989⋆
RT-GARCH	0.1742 (6)	0.0014	0.2558 (4)	0.3299⋆	0.3609 (3)	0.0583⋆	0.4340 (3)	0.0989⋆
RT-GARCH-L	0.1763 (9)	0.0014	0.2663 (10)	0.3299⋆	0.3857 (9)	0.0027	0.4640 (8)	0.0016
RT-GARCH-LF	0.1695 (3)	0.9936⋆	0.2591 (7)	0.3299⋆	0.3835 (8)	0.0000	0.4724 (10)	0.0028
ERT-GARCH	0.1719 (4)	0.0014	0.2558 (3)	0.2259⋆	0.3626 (4)	0.1110⋆	0.4362 (4)	0.0989⋆
ERT-GARCH-L	0.1772 (10)	0.0014	0.2593 (8)	0.3299⋆	0.3633 (5)	0.0583⋆	0.4294 (2)	0.0740⋆
ERT-GARCH-LF	<u>0.1545</u> (2)	0.0536⋆	<u>0.2464</u> (1)	0.0881⋆	<u>0.3357</u> (1)	0.0583⋆	<u>0.4142</u> (1)	0.0740⋆
SV-Kalman	0.2278 (11)	0.0140	0.2978 (11)	1.0000⋆	0.3934 (10)	0.0027	0.4678 (9)	0.0028
SV-Bellman	0.1746 (7)	1.0000⋆	0.2651 (9)	0.6284⋆	0.4199 (11)	0.0027	0.5704 (11)	0.0028
Stoxx 50								
GARCH	0.1992 (6)	0.0311	0.2657 (5)	0.4948⋆	0.3839 (2)	0.0543⋆	0.4533 (4)	0.7242⋆
E-GARCH	0.1972 (4)	0.0311	0.2643 (4)	0.4948⋆	0.3852 (3)	0.0939⋆	0.4551 (6)	0.7242⋆
ET-GARCH	0.1790 (3)	0.5967⋆	0.2635 (3)	0.1531⋆	0.4051 (7)	0.8831⋆	0.4721 (8)	0.0002
RT-GARCH	0.2073 (7)	0.0311	0.2806 (7)	0.5166⋆	0.3941 (4)	0.1457⋆	0.4544 (5)	1.0000⋆
RT-GARCH-L	0.2193 (10)	0.0311	0.2964 (10)	0.8565⋆	0.4110 (8)	0.8831⋆	0.4743 (9)	0.0102
RT-GARCH-LF	0.1984 (5)	1.0000⋆	0.2777 (6)	0.4948⋆	0.3949 (5)	0.8831⋆	0.4673 (7)	0.0050
ERT-GARCH	0.1719 (1)	0.0311	<u>0.2558</u> (1)	0.4948⋆	<u>0.3626</u> (1)	0.8831⋆	0.4362 (1)	0.0000
ERT-GARCH-L	0.2101 (8)	0.0311	0.2831 (8)	0.8565⋆	0.4158 (9)	1.0000⋆	0.5143 (10)	0.0102
ERT-GARCH-LF	<u>0.1772</u> (2)	0.2446⋆	0.2599 (2)	0.1531⋆	0.4690 (11)	0.0127	<u>0.4443</u> (2)	0.2250⋆
SV-Kalman	0.2713 (11)	0.0471	0.3213 (11)	1.0000⋆	0.3972 (6)	0.1457⋆	0.4478 (3)	0.5557⋆
SV-Bellman	0.2107 (9)	0.0311	0.2874 (9)	0.6151⋆	0.4228 (10)	0.0000	0.5367 (11)	0.0102

Note: p_{MCS} are the p-values from the Model Confidence Set test of Hansen et al. (2011). The ones marked with an ⋆ mean that the corresponding model is in $\hat{\mathcal{M}}_{0.95}$. Underlined values correspond to the model with the smallest MSPE that is also in the MCS. The MSPE rank from small to large is given in parentheses.

4 Conclusion

The contributory volatility forecasts from the Real-Time GARCH model by Smetanina (2017a) are spurious as a result of incorrect proofs. We correct these proofs and now find that the RT-GARCH forecasts are not convincingly better than those of conventional GARCH models, but better than SV models. In an attempt to improve on RT-GARCH, we introduce the new Enhanced RT-GARCH model, which better captures asymmetric effects of returns on volatility. We conclude that the ERT-GARCH model yields superior volatility forecasts compared to RT-GARCH, conventional GARCH and stochastic volatility models.

In an empirical study we first estimate the aforementioned models in a full-sample analysis for two global stock indices. The current return innovation in RT-GARCH and ERT-GARCH models results in a better in-sample fit compared to the conventional GARCH and SV models. Furthermore, we find that the RT-GARCH-LF model, which captures leverage and feedback effects, might falsely be selected as the best model. This is caused by a subtle misspecification in Smetanina (2017a), leading to an artificially increased log-likelihood. Overall, ERT-GARCH-LF is selected as the best volatility model.

Subsequently we estimate all models within a rolling window procedure to create 1-, 5-, 15- and 30- day ahead volatility forecasts. None of the models provide unbiased or efficient forecasts, even though the short-term forecasts (1- and 5-day ahead) can explain a large part of the variance in realized volatility. Considering forecast accuracy based on the MSPE loss function, two models really stand out: ERT-GARCH and ERT-GARCH-LF. Consistent with the full-sample analysis, we generally find that ERT-GARCH-LF is the most accurate, but in one specific case it performs the worst out of all models. We therefore conclude that one should always use ERT-GARCH with- and without leverage and feedback effects, to see which model best fits the asset and forecast horizon.

On the opposite side we observe that the SV model based on the Kalman filter produces the least accurate short-term forecasts. The least accurate long-term forecasts (15- and 30-day ahead) are obtained by the SV model based on the recently developed Bellman filter (see Lange, 2020). The poor accuracy is as expected, because the SV models have the least number of parameters and do not incorporate leverage or feedback effects. For further research it would hence be interesting to add the asymmetric SV model by Catania (2020) into the model comparison.

To really be able to call ERT-GARCH better than RT-GARCH, GARCH and SV models, a broader generalization of our results is needed. For further research we thus suggest to not only make our comparison for stock indices, but also for individual stocks and foreign exchange indices. Moreover, it would be interesting to see whether our improved volatility forecasts excel in practical applications in the fields of risk management or asset allocation. Finally, Student's t-distributed innovations and included realized measures might further improve empirical performance, as already suggested by Smetanina (2017a).

References

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* (Vol. 55). Washington, DC: National Bureau of Standards.

Andersen, T., Chung, H.-J., & Sorensen, B.E., (1997). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics, 91*, 61–87.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics, 61(1)*, 43-76.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association*, 171–181.

Bollerslev, T. (1986). Generalized autoregressive heteroskedasticity. *Journal of Econometrics, 31*, 307–327.

Catania, L. (2020). A stochastic volatility model with a general leverage specification. *Journal of Business & Economic Statistics*, 1-23.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica, 41(1)*, 135-155.

Ding, Y. (2020). Weak diffusion limit of real-time GARCH models: The role of current return information. *Cambridge Working Paper in Economics 20112.*

Ding, Y. (2021). Augmented real-time GARCH: A joint model for returns, volatility and volatility of volatility. *Cambridge Working Papers in Economics 2112*, Faculty of Economics, University of Cambridge.

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

Ederington, L. H. & Guan, W. (2010). The bias in time series volatility forecasts. *The Journal of Futures Markets, 30(4)*, 305–323.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*, 987–1007.

Francq, C., & Zakoian, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications, 117(9)*, 1265-1284.

Fridman, M., & Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business & Economic Statistics, 16(3)*, 284-291.

Hansen, P. R., & A. Lunde. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1,1)? *Journal of Applied Econometrics, 20*, 873–889.

Hansen P. R., Lunde, A., & Nason J. M. (2011). The model confidence set. *Econometrica 79*, 453–497.

Harvey, A. C. (1981). *Time Series Models*. New York: John Wiley.

Harvey. A.C., Ruiz, E. & Shephard, N.G. (1992). Multivariate stochastic variance models. *The Review of Economic Studies, 61(2)*, 247-264.

Jacquier, E., Polson, N.G., & Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics, 12(4)*, 371–389.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *American Society of Mechanical Engineers: Journal of Basic Engineering, 82(1)*, 35-45.

Lange, R.-J. (2020). Bellman filtering for state-space models. *Tinbergen Institute Discussion Paper, 2020-052(3).*

Lange, R.-J. (2021). *A short explanation of RT-GARCH (with proofs).* Unpublished article. Department of Econometrics, Erasmus University Rotterdam.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394-419.

Nelson, D. B. (1988). *The time series behavior of stock market volatility and returns.* Doctoral dissertation, Massachusetts Institute of Technology.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*, 347–370.

Ng, H., & Lam, K.-P. (2006). How does the sample size affect GARCH models? *Proceedings of the 2006 Joint Conference on Information Sciences.* Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong.

Patton, A. (2011). Volatility forecast evaluation and comparison using imperfect volatility proxies. *Journal of Econometrics 160*, 246–256.

Smetanina, E. (2017a). Real-Time GARCH. *Journal of Financial Econometrics, 15 (4)*, 561–601.

Smetanina, E. (2017b). *Supplementary material for "Real-Time GARCH: Does current information matter?"*. Unpublished article. University of Cambridge. Retrieved from https://sites.google.com/site/smetaninakatja/research.

Smetanina, E., & Wu, W. B. (2019). *Asymptotic theory for QMLE for the real-time GARCH(1,1) model.* Working Paper, University of Chicago. Retrieved from https://sites.google.com/site/smetaninakatja/research.

Taylor, S. (1986). *Modeling Financial Time Series.* New York: John Wiley and Sons.

Wooldridge, J. M. (1991). On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics, 47(1)*, 5-46.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control, 18*, 931–955.

A Proofs

A.1 Proof Result 1

We now derive the unconditional moment. Combining (2) and (3) gives:

$$\lambda_t^2 = \alpha + \beta\lambda_{t-1}^2 + \left(\gamma_1 \mathbb{1}_{r_t \leq 0} + \gamma_2 \mathbb{1}_{r_t > 0}\right) r_{t-1}^2 + \left(\phi_1 \mathbb{1}_{z_t \leq 0} + \phi_2 \mathbb{1}_{z_t > 0}\right) z_t^2. \tag{21}$$

As the model assumes z_t has mean zero and a symmetric distribution, z_t^2 is independent from both $\mathbb{1}_{z_t \leq 0}$ and $\mathbb{1}_{z_t > 0}$. Taking the expectation of (2) gives:

$$\begin{aligned} E[\lambda_t^2] &= E[\sigma_t^2] + \left(\phi_1 E[\mathbb{1}_{z_t \leq 0}] + \phi_2 E[\mathbb{1}_{z_t > 0}]\right) E[z_t^2] \\ &= E[\sigma_t^2] + \left(\phi_1 P[z_t \leq 0] + \phi_2 P[z_t > 0]\right) E[z_t^2] \\ &= E[\sigma_t^2] + \frac{1}{2}(\phi_1 + \phi_2) \\ &= E[\sigma_t^2] + \phi. \end{aligned} \tag{22}$$

We now write the following term:

$$
\begin{aligned}
\mathbb{1}_{r_t \leq 0} r_{t-1}^2 &= \mathbb{1}_{z_t \leq 0} \cdot \lambda_{t-1}^2 z_{t-1}^2 \\
&= \mathbb{1}_{z_t \leq 0} \cdot \left(\sigma_{t-1}^2 + \phi_1 z_{t-1}^2 \mathbb{1}_{z_{t-1} \leq 0} + \phi_2 z_{t-1}^2 \mathbb{1}_{z_{t-1} > 0} \right) z_{t-1}^2 .
\end{aligned}
$$

Combined with equation (22), taking the expectation of this term yields:

$$
\begin{aligned}
E[\mathbb{1}_{r_t \leq 0} r_{t-1}^2] &= \frac{1}{2} E[\sigma_{t-1}^2] + \frac{1}{4} \phi_1 E[z_{t-1}^4] + \frac{1}{4} \phi_2 E[z_{t-1}^4] \\
&= \frac{1}{2} \left(E[\lambda_{t-1}^2] - \phi + \frac{1}{2} E[z_{t-1}^4](\phi_1 + \phi_2) \right) \\
&= \frac{1}{2} \left(E[\lambda_{t-1}^2] + \ \phi(E[z_{t-1}^4] - 1) \right) . \qquad (23)
\end{aligned}
$$

Similarly, it can be shown that $E[\mathbb{1}_{r_t > 0} r_{t-1}^2]$ is also equal to (23). Now taking the expectation of equation (21) results in:

$$
\begin{aligned}
E[\lambda_t^2] &= \alpha + \beta E[\lambda_{t-1}^2] + \frac{1}{2}(\gamma_1 + \gamma_2) \left(E[\lambda_{t-1}^2] + \ \phi(E[z_{t-1}^4] - 1) \right) + \phi \\
&= \alpha + \beta E[\lambda_{t-1}^2] + \gamma \left(E[\lambda_{t-1}^2] + \ \phi(E[z_{t-1}^4] - 1) \right) + \phi .
\end{aligned}
$$

Assuming $E[\lambda_t^2] = E[\lambda_{t-1}^2]$ and solving for $E[\lambda_t^2]$ gives the result:

$$
\begin{aligned}
E[\lambda_t^2] &= \alpha + \beta E[\lambda_t^2] + \gamma E[\lambda_t^2] + \gamma \phi (E[z_{t-1}^4] - 1) + \phi \\
&= \frac{\alpha + \phi + \gamma \phi (E[z_{t-1}^4] - 1)}{1 - \gamma - \beta} . \qquad (24)
\end{aligned}
$$

These expressions are valid long as $\beta + \gamma < 1$. We may initialize the process λ_t^2 at time zero using the unconditional distribution: $\lambda_0^2 = E[\lambda_t^2]$ as given above.

A.2 Proof Result 4

The expression for one-step ahead volatility forecasts are obtained directly from equation (2):

$$
\begin{aligned}
E_t[\lambda_{t+1}^2] &= E_t[\sigma_{t+1}^2] + E_t[\phi_1 z_{t+1}^2 \mathbb{1}_{z_{t+1} \leq 0} + \phi_2 z_{t+1}^2 \mathbb{1}_{z_{t+1} > 0}] \\
&= \alpha + \beta \lambda_t^2 + \gamma_1 E_t[\mathbb{1}_{r_{t+1} \leq 0}] r_t^2 + \gamma_2 E_t[\mathbb{1}_{r_{t+1} > 0}] r_t^2 + \phi \\
&= \alpha + \beta \lambda_t^2 + \gamma r_t^2 + \phi . \qquad (25)
\end{aligned}
$$

In order to derive the k-step ahead forecasts, we first write α in equation (24) in terms of $E[\lambda_t^2] = \lambda^2$ and $E[z_{t-1}^4] = K_z$:

$$\alpha = \lambda^2(1 - \gamma - \beta) - \phi - \gamma\phi(K_z - 1).$$

The process for λ_t^2, given in equation (2), can now be written as:

$$\begin{aligned}
\lambda_{t+1}^2 &= \alpha + \beta\lambda_t^2 + \gamma_1 r_t^2 \mathbb{1}_{r_{t+1}\le 0} + \gamma_2 r_t^2 \mathbb{1}_{r_{t+1}>0} + \phi_1 z_{t+1}^2 \mathbb{1}_{z_{t+1}\le 0} + \phi_2 z_{t+1}^2 \mathbb{1}_{z_{t+1}>0} \\
&= \alpha + \beta\lambda_t^2 + \gamma_1 \lambda_t^2 z_t^2 \mathbb{1}_{z_{t+1}\le 0} + \gamma_2 \lambda_t^2 z_t^2 \mathbb{1}_{z_{t+1}>0} + \phi_1 z_{t+1}^2 \mathbb{1}_{z_{t+1}\le 0} + \phi_2 z_{t+1}^2 \mathbb{1}_{z_{t+1}>0} \\
&= \alpha + \phi + (\beta + \gamma)\lambda_t^2 + \gamma_1 \lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) + \gamma_2 \lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}>0} - \frac{1}{2} \right) \\
&\quad + \phi_1 \left(z_{t+1}^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) + \phi_2 \left(z_{t+1}^2 \mathbb{1}_{z_{t+1}>0} - \frac{1}{2} \right) \\
&= \lambda^2(1 - \gamma - \beta) - \gamma\phi(K_z - 1) + (\beta + \gamma)\lambda_t^2 + \gamma_1 \lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) + \gamma_2 \lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}>0} - \frac{1}{2} \right) \\
&\quad + \phi_1 \left(z_{t+1}^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) + \phi_2 \left(z_{t+1}^2 \mathbb{1}_{z_{t+1}>0} - \frac{1}{2} \right) \\
&= \lambda^2(1 - \gamma - \beta) + (\beta + \gamma)\lambda_t^2 + \gamma_1 \underbrace{\left\{ \lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) - \frac{1}{2}\phi(K_z - 1) \right\}}_{\xi_{1t}} \\
&\quad + \gamma_2 \underbrace{\left\{ \lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}>0} - \frac{1}{2} \right) - \frac{1}{2}\phi(K_z - 1) \right\}}_{\xi_{2t}} + \phi_1 \underbrace{\left(z_{t+1}^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right)}_{\zeta_{1t}} + \phi_2 \underbrace{\left(z_{t+1}^2 \mathbb{1}_{z_{t+1}>0} - \frac{1}{2} \right)}_{\zeta_{2t}}.
\end{aligned} \tag{26}$$

We can derive that $E_{t-1}[\xi_{1t}] = 0$. This can be seen by computing

$$\begin{aligned}
E_{t-1}\left[\lambda_t^2 \left(z_t^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) \right] &= E_{t-1}\left[\left(\sigma_t^2 + (\phi_1 \mathbb{1}_{z_t\le 0} + \phi_2 \mathbb{1}_{z_t>0})\, z_t^2 \right) \left(z_t^2 \mathbb{1}_{z_{t+1}\le 0} - \frac{1}{2} \right) \right] \\
&= \left(\frac{1}{2}\sigma_t^2 + \frac{1}{2}\phi K_z \right) - \left(\frac{1}{2}\sigma_t^2 + \frac{1}{2}\phi \right) \\
&= \frac{1}{2}\phi(K_z - 1).
\end{aligned}$$

Similarly, it can be shown that $E_{t-1}[\xi_{2t}] = 0$. To see that $E_{t-1}[\zeta_{1t}] = 0$ and $E_{t-1}[\zeta_{2t}] = 0$ we compute:

$$E_{t-1}\left[z_{t+1}^2 \mathbb{1}_{z_{t+1}\le 0} \right] = \frac{1}{2} = E_{t-1}\left[z_{t+1}^2 \mathbb{1}_{z_{t+1}>0} \right].$$

Accordingly, we obtain from equation (26) that

$$\begin{aligned} E_t[\lambda^2_{t+k}] &= \lambda^2(1-\gamma-\beta) + (\beta+\gamma)E_t[\lambda^2_{t+k-1}] \\ &= \lambda^2 - (\beta+\gamma)\lambda^2 + (\beta+\gamma)E_t[\lambda^2_{t+k-1}] \\ &= \lambda^2 + (\beta+\gamma)E_t[\lambda^2_{t+k-1} - \lambda^2], \qquad k \geq 2. \end{aligned} \tag{27}$$

It is important that $k \geq 2$ to ensure $E_t[\xi_{t+k-1}] = 0$, as z_t is known at time t and thus $E_t[\xi_{it}] \neq 0$ for $i = 1, 2$.

Summing up the derivations, we have found:

$$\begin{aligned} E_t[\lambda^2_{t+1}] &= \alpha + \beta\lambda^2_t + \gamma r^2_t + \phi, \\ E_t[\lambda^2_{t+2}] &= \lambda^2 + (\beta+\gamma)E_t[\lambda^2_{t+1} - \lambda^2], \\ E_t[\lambda^2_{t+3}] &= \lambda^2 + (\beta+\gamma)E_t[\lambda^2_{t+2} - \lambda^2] = \lambda^2 + (\beta+\gamma)^2 E_t[\lambda^2_{t+1} - \lambda^2], \\ &\vdots \\ E_t[\lambda^2_{t+k}] &= \lambda^2 + (\beta+\gamma)E_t[\lambda^2_{t+k-1} - \lambda^2] = \lambda^2 + (\beta+\gamma)^{k-1} E_t[\lambda^2_{t+1} - \lambda^2]. \end{aligned}$$

The last line implies:

$$\begin{aligned} E_t[\lambda^2_{t+k}] &= \lambda^2 + (\beta+\gamma)^{k-1} E_t[\lambda^2_{t+1} - \lambda^2] \\ &= \lambda^2 + (\beta+\gamma)^{k-1}(\alpha + \beta\lambda^2_t + \gamma r^2_t + \phi - \lambda^2), \qquad k \geq 1. \end{aligned} \tag{28}$$

Now compute the return variance forecast using relation (22):

$$\begin{aligned} E_t[r^2_{t+k}] &= E_t[\lambda^2_{t+k} z^2_{t+k}] \\ &= E_t[(\sigma^2_{t+k} + \phi_1 z^2_{t+k} \mathbb{1}_{z_{t+k} \leq 0} + \phi_2 z^2_{t+k} \mathbb{1}_{z_{t+k} > 0}) z^2_{t+k}] \\ &= E_t[\sigma^2_{t+k}] + \phi K_z \\ &= E_t[\lambda^2_{t+k}] + \phi(K_z - 1), \qquad k \geq 1. \end{aligned} \tag{29}$$

Combining equations (28) and (29) gives the final result:

$$E_t[r^2_{t+k}] = \lambda^2 + (\beta+\gamma)^{k-1}(\alpha + \beta\lambda^2_t + \gamma r^2_t + \phi - \lambda^2) + \phi(K_z - 1). \tag{30}$$

A.3 Proof Result 10

Having found the model parameter estimates, we can find the expression for the k-step ahead forecast of h_t. The 1-step ahead forecast of h_t is

$$\begin{aligned} E_t[h_{t+1}] &= E_t[\omega + \delta h_t + \eta_{t+1}] \\ &= \omega + \delta h_t. \end{aligned}$$

The 2-step ahead forecast of h_t is

$$\begin{aligned} E_t[h_{t+2}] &= E_t[\omega + \delta h_{t+1} + \eta_{t+2}] \\ &= \omega + \delta E_t[h_{t+1}] \\ &= \omega + \delta(\omega + \delta h_t) \\ &= \omega + \delta\omega + \delta^2 h_t. \end{aligned}$$

The 3-step ahead forecast of h_t is

$$\begin{aligned} E_t[h_{t+3}] &= E_t[\omega + \delta h_{t+2} + \eta_{t+3}] \\ &= \omega + \delta E_t[h_{t+2}] \\ &= \omega + \delta(\omega + \delta\omega + \delta^2 h_t) \\ &= \omega + \delta\omega + \delta^2\omega + \delta^3 h_t. \end{aligned}$$

Recursively we find the expression for the k-step ahead forecast of h_t:

$$\begin{aligned} E_t[h_{t+k}] &= \omega(1 + \delta + \delta^2 + ... + \delta^{k-1}) + \delta^k h_t \\ &= \omega\left(\frac{1-\delta^k}{1-\delta}\right) + \delta^k h_t. \end{aligned} \tag{31}$$

The 1-step ahead forecasting error of h_t is:

$$\begin{aligned} h_{t+1} - h_{t+1|t} &= \omega + \delta h_t + \eta_{t+1} - \omega - \delta h_t \\ &= \eta_{t+1}. \end{aligned}$$

The 2-step ahead forecasting error of h_t is:

$$\begin{aligned}
h_{t+2} - h_{t+2|t} &= \omega + \delta h_{t+1} + \eta_{t+2} - \omega - \delta h_{t+1|t} \\
&= \eta_{t+2} + \delta(h_{t+1} - h_{t+1|t}) \\
&= \eta_{t+2} + \delta \eta_{t+1}.
\end{aligned}$$

The 3-step ahead forecasting error of h_t is:

$$\begin{aligned}
h_{t+3} - h_{t+3|t} &= \omega + \delta h_{t+2} + \eta_{t+3} - \omega - \delta h_{t+2|t} \\
&= \eta_{t+3} + \delta(h_{t+2} - h_{t+2|t}) \\
&= \eta_{t+3} + \delta(\eta_{t+2} + \delta \eta_{t+1}) \\
&= \eta_{t+3} + \delta \eta_{t+2} + \delta^2 \eta_{t+1}.
\end{aligned}$$

Recursively we find the k-step ahead forecasting error of h_t:

$$h_{t+k} - h_{t+k|t} = \eta_{t+k} + \delta h_{t+k-1} + \cdots + \delta^{k-1} \eta_{t+1}.$$

The k-step ahead variance forecast of h_t can be found using the fact that $\text{Var}_t[h_{t+k}] = E[(h_{t+k} - h_{t+k|t})^2]$:

$$\begin{aligned}
\text{Var}_{h_{t+1}} &= E_t[(\eta_{t+1})^2] = \sigma_\eta^2 \\
\text{Var}_{h_{t+2}} &= E_t[(\eta_{t+2} + \delta \eta_{t+1})^2] = \sigma_\eta^2 + \sigma_\eta^2 \delta^2 \\
\text{Var}_{h_{t+3}} &= E_t[(\eta_{t+3} + \delta \eta_{t+2} + \delta^2 \eta_{t+1})^2] = \sigma_\eta^2 + \sigma_\eta^2 \delta^2 + \sigma_\eta^2 \delta^4 \\
&\vdots \\
\text{Var}_{h_{t+k}} &= E_t[(\eta_{t+k} + \delta h_{t+k-1} + \cdots + \delta^{k-1} \eta_{t+1})^2] \\
&= \sigma_\eta^2 (1 + \delta^2 + \delta^4 + \delta^{k-1}) \\
&= \sigma_\eta^2 \left(\frac{1 - \delta^{2k}}{1 - \delta^2} \right). \qquad (32)
\end{aligned}$$

From the state equation (15) of the SV_{NL} model and using that $\exp(h_t)$ is log-normally distributed, we find that the k-step ahead forecast of r^2_{t+k} can be written as:

$$\begin{aligned} E_t[r^2_{t+k}] &= E_t\left[(\epsilon_t \exp(h_{t+k}/2))^2\right] \\ &= E_t\left[\epsilon_t^2\right] E_t\left[\ \exp(h_{t+k})\right] \\ &= \exp\left(E_t[h_{t+k}] + \frac{1}{2}\mathrm{Var}_t[h_{t+k}]\right), \end{aligned} \tag{33}$$

where $E_t[h_{t+k}]$ and $\mathrm{Var}_t[h_{t+k}]$ are given by equations (31) and (32) respectively. Finally, the volatility prediction is the square root of $E_t[r^2_{t+k}]$.

B Full-Sample Analysis

Table 6: Parameter estimates, AIC and log-likelihood of the symmetric and asymmetric GARCH models for S&P 500 and Stoxx 50 over the period 2 January 2004 - 31 March 2021. Numerically calculated standard errors are in parentheses.

Returns	μ	α	β	γ_1	γ_2	AIC	L_T
S&P 500							
GARCH	-	0.0185***	0.8544***	0.1265***	-	2.4145	-5222
		(0.0032)	(0.0095)	(0.0097)			
E-GARCH	0.0422***	0.0188***	0.8523***	0.1285***	-	2.4135	-5219
	(0.0101)	(0.0024)	(0.0096)	(0.0098)			
ET-GARCH	0.0140*	0.0190***	0.8809***	0.1846***	0.0000	2.3833	-5153
	(0.0102)	(0.0021)	(0.0082)	(0.0139)	(0.0088)		
Stoxx 50							
GARCH	-	0.0246***	0.8773***	0.1047***	-	2.8251	-6215
		(0.0034)	(0.0083)	(0.0082)			
E-GARCH	0.0401***	0.0254***	0.8736***	0.1080***	-	2.8226	-6209
	(0.0129)	(0.0035)	(0.0086)	(0.0085)			
ET-GARCH	0.0054	0.0236***	0.8900***	0.1798***	0.0000	2.7751	-6103
	(0.0125)	(0.0028)	(0.0076)	(0.0137)	(0.0072)		

Note: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7: Parameter estimates, AIC and log-likelihood of the SV models for S&P 500 and Stoxx 50 over the period 2 January 2004 - 31 March 2021. Numerically calculated standard errors are in parentheses.

Returns	ω	δ	σ_η^2	AIC	L_T
S&P 500					
SV-Kalman	0.0162***	0.9782***	0.0560***	4.6126	-9979
	(0.0045)	(0.0043)	(0.0087)		
SV-Bellman	0.0006	0.9712***	0.0732***	2.3607	-5106
	(0.0050)	(0.0047)	(0.0105)		
Stoxx 50					
SV-Kalman	-0.0073**	0.9769***	0.0381***	4.5733	-10063
	(0.0032)	(0.0051)	(0.0082)		
SV-Bellman	0.0087**	0.9626***	0.0713***	2.7741	-6103
	(0.0044)	(0.0064)	(0.0119)		

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C Out-of-Sample Forecast Evaluation

Table 8: Parameter estimates and the R^2 of the Mincer-Zarnowitz regression RV_{t+30} on $\hat{\sigma}_{t+30|t}$ at the 1-, 5-, 15- and 30-day ahead horizons. The standard errors are given in parentheses.

	S&P 500					Stoxx 50				
	α_0	S.E.	α_1	S.E.	R^2	α_0	S.E.	α_1	S.E.	R^2
$k = 1$										
GARCH	-0.046***	(0.013)	0.929***	(0.011)	0.676	-0.036**	(0.018)	0.955***	(0.009)	0.588
E-GARCH	-0.047***	(0.013)	0.929***	(0.011)	0.679	-0.029*	(0.017)	0.949***	(0.010)	0.592
ET-GARCH	-0.040***	(0.012)	0.921***	(0.011)	0.718	0.067***	(0.015)	0.876***	(0.011)	0.638
RT-GARCH	-0.070***	(0.013)	0.968***	(0.010)	0.667	-0.196***	(0.020)	1.105***	(0.008)	0.570
RT-GARCH-L	-0.191***	(0.015)	1.090***	(0.009)	0.669	-0.376***	(0.023)	1.269***	(0.007)	0.563
RT-GARCH-LF	-0.153***	(0.014)	1.049***	(0.010)	0.680	-0.358***	(0.021)	1.254***	(0.007)	0.608
ERT-GARCH	-0.076***	(0.013)	0.973***	(0.010)	0.672	-0.193***	(0.020)	1.101***	(0.008)	0.577
ERT-GARCH-L	-0.090***	(0.014)	0.988	(0.010)	0.661	-0.300***	(0.022)	1.202***	(0.007)	0.574
ERT-GARCH-LF	-0.093***	(0.013)	1.017**	(0.010)	0.698	-0.016	(0.016)	0.961***	(0.010)	0.626
SV-Kalman	0.064***	(0.015)	0.924***	(0.010)	0.540	-0.054**	(0.024)	1.126***	(0.007)	0.430
SV-Bellman	-0.005	(0.013)	0.951***	(0.010)	0.652	-0.076***	(0.020)	1.068***	(0.008)	0.547
$k = 5$										
GARCH	-0.014	(0.016)	0.875***	(0.010)	0.527	0.010	(0.022)	0.907***	(0.009)	0.451
E-GARCH	-0.016	(0.016)	0.875***	(0.010)	0.528	0.016***	(0.022)	0.902***	(0.009)	0.455
ET-GARCH	0.001	(0.016)	0.862***	(0.010)	0.537	0.128***	(0.019)	0.815***	(0.010)	0.470
RT-GARCH	-0.049***	(0.018)	0.932***	(0.009)	0.509	-0.210***	(0.028)	1.110***	(0.007)	0.413
RT-GARCH-L	-0.204***	(0.02)	1.082***	(0.008)	0.493	-0.423***	(0.033)	1.290***	(0.006)	0.398
RT-GARCH-LF	-0.174***	(0.019)	1.039***	(0.008)	0.512	-0.405***	(0.03)	1.273***	(0.006)	0.439
ERT-GARCH	-0.058***	(0.018)	0.938***	(0.009)	0.510	-0.219***	(0.028)	1.116***	(0.007)	0.418
ERT-GARCH-L	-0.060***	(0.018)	0.946***	(0.009)	0.499	-0.371***	(0.031)	1.267***	(0.006)	0.418
ERT-GARCH-LF	-0.034**	(0.017)	0.935***	(0.009)	0.520	-0.046**	(0.023)	0.982**	(0.008)	0.448
SV-Kalman	0.093***	(0.018)	0.851***	(0.009)	0.410	-0.003	(0.029)	1.043***	(0.006)	0.310
SV-Bellman	-0.015	(0.017)	0.887***	(0.010)	0.500	-0.077***	(0.026)	1.003	(0.007)	0.391
$k = 15$										
GARCH	0.042*	(0.022)	0.793***	(0.009)	0.335	0.170***	(0.031)	0.758***	(0.007)	0.220
E-GARCH	0.039*	(0.022)	0.793***	(0.009)	0.335	0.176***	(0.03)	0.751***	(0.007)	0.220
ET-GARCH	0.042*	(0.022)	0.803***	(0.009)	0.337	0.304***	(0.028)	0.650***	(0.008)	0.207
RT-GARCH	0.025	(0.024)	0.846***	(0.008)	0.307	-0.013	(0.043)	0.937***	(0.005)	0.165
RT-GARCH-L	-0.130***	(0.031)	0.994	(0.006)	0.253	-0.184***	(0.055)	1.065***	(0.004)	0.139
RT-GARCH-LF	-0.138***	(0.030)	0.972***	(0.007)	0.277	-0.246***	(0.050)	1.114***	(0.004)	0.177
ERT-GARCH	0.017	(0.024)	0.852***	(0.008)	0.304	-0.022	(0.044)	0.943***	(0.005)	0.163
ERT-GARCH-L	0.027	(0.024)	0.849***	(0.008)	0.299	0.034	(0.049)	0.889***	(0.005)	0.123
ERT-GARCH-LF	-0.088***	(0.025)	1.024***	(0.007)	0.327	0.519***	(0.045)	0.482***	(0.005)	0.042
SV-Kalman	0.162***	(0.023)	0.729***	(0.008)	0.251	0.143***	(0.040)	0.857***	(0.005)	0.146
SV-Bellman	0.007	(0.024)	0.763***	(0.009)	0.304	0.062	(0.039)	0.799***	(0.006)	0.174
$k = 30$										
GARCH	0.126***	(0.028)	0.696***	(0.008)	0.194	0.333***	(0.040)	0.610***	(0.006)	0.094
E-GARCH	0.123***	(0.028)	0.696***	(0.008)	0.193	0.337***	(0.040)	0.605***	(0.006)	0.094
ET-GARCH	0.112***	(0.029)	0.727***	(0.007)	0.189	0.452***	(0.037)	0.514***	(0.007)	0.082
RT-GARCH	0.139***	(0.029)	0.729***	(0.007)	0.171	0.341***	(0.057)	0.630***	(0.004)	0.047
RT-GARCH-L	0.092**	(0.041)	0.773***	(0.005)	0.099	0.453***	(0.072)	0.515***	(0.003)	0.021
RT-GARCH-LF	0.067**	(0.040)	0.763***	(0.005)	0.111	0.323***	(0.068)	0.622***	(0.004)	0.035
ERT-GARCH	0.134***	(0.030)	0.732***	(0.007)	0.167	0.352***	(0.058)	0.620***	(0.004)	0.044
ERT-GARCH-L	0.078**	(0.036)	0.857***	(0.005)	0.136	0.914***	(0.059)	0.120***	(0.004)	0.002
ERT-GARCH-LF	0.062*	(0.032)	0.856***	(0.006)	0.173	0.266***	(0.058)	0.712***	(0.004)	0.055
SV-Kalman	0.260***	(0.027)	0.600***	(0.008)	0.147	0.357***	(0.051)	0.635***	(0.004)	0.055
SV-Bellman	0.124***	(0.030)	0.600***	(0.008)	0.168	0.358***	(0.049)	0.521***	(0.006)	0.059

Note: reject H_0: $\alpha_0 = 0$ with $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$, reject H_0: $\alpha_1 = 1$ with $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

D Brief Explanation of Matlab Code

Map 1 is used to estimate the RT-GARCH models from Smetanina (2017) and obtain their coefficients, Standard Errors, log-likelihood and AIC that are presented in Table 3 of the paper. The "_se" functions are the non-nested variants of the general RT-GARCH_gaussian function, which are necessary to calculate the finite Hessian in "fdhess".

Map 2 is used to estimate the new ERT-GARCH models and obtain their coefficients, Standard Errors, log-likelihood and AIC that are presented in Table 3 of the paper. The "_se" functions are the non-nested variants of the general ERT-GARCH_gaussian function, which are necessary to calculate the finite Hessian in "fdhess".
Note that there is an additional model that can be estimated: the "MUZERO-LF". This corresponds to the quasi-deterministic RT-GARCH model as mentioned in Section 3.2 of the paper.

Map 3 is used to estimate the SV models using the Kalman and Bellman filter and obtain their coefficients, Standard Errors, log-likelihood and AIC that are presented in Table 7 of the paper. The functions "KalmanFilter" and "NegativeLikelihood" are used to estimate via the Kalman filter. The functions "Bellman_filter", "info", "info_expected", "link", "link_inverse", "link2", "logpdf", "NegativeLoglikelihood" and "score" are codes from Lange (2020) used to estimate via the Bellman filter.

Map 4 is used to estimate and forecast all models in a rolling window procedure. The same estimation procedures as in map 1 to 3 apply, but now the returns sample changes every iteration. The workfiles are saved in the folder "aaa_Resultaten". The results for the RT-GARCH-LF model of SP 500 are for example saved as "sp_rtgarchlf.mat".

Map 5 is used to obtain the out-of-sample forecast evaluation results. The "aa_Main_Forecast_Evaluation" script picks the rolling window workfiles and calculates the results for unbiasedness, efficiency and accuracy. Also, VaR forecasts are made and their statistics for conditional and unconditional coverage are calculated (note: these are not used in the paper). The "aa_Relative_Accuracy" script calculates the MCS from Hansen et al. (2011), using the functions "mcs" and "bootstrap_block" obtained from the MFE Toolbox (mfe-toolbox/bootstrap at master · bashtage/mfe-toolbox · GitHub).