

A Family of Stochastic Unit GARCH Models

Mamadou Abdoulaye Konté

Correspondence: Mamadou Abdoulaye Konté, BP 234, UFR SEG, Gaston Berger University, Senegal. Tel: 221-77-622-1350. E-mail: kontedoudou@yahoo.fr or mamadou-abdoulaye.konte@ugb.edu.sn

Received: September 29, 2012 Accepted: November 28, 2012 Online Published: December 10, 2012

doi:10.5539/ijef.v5n1p177 URL: <http://dx.doi.org/10.5539/ijef.v5n1p177>

Abstract

A class of Asymmetric GARCH models is presented. It shares the same unconditional variance and volatility forecast formula as the standard GARCH(P,Q) model under the assumption of a symmetric conditional distribution for innovations. We use three models of this class to assess their ability to forecast S&P 500 market volatility and to make better decisions for the purpose of risk management and investment. Subsequently, a comparison is made with respect to competing models (GARCH, EGARCH, GJR). It was found that for the in-sample evaluation, the best model is obtained from the Stochastic Unit GARCH (SUGARCH) model where leverage effects are introduced through the GARCH (i.e. β_1) parameter. For the out-of-sample evaluation (QLIKE loss function), it is better to use the SUGARCH class where the asymmetry appears on the ARCH (i.e. α_1) parameter.

Keywords: time-varying coefficients, GARCH models, market volatility, forecasting

1. Introduction

Understanding how the market volatility evolves is challenging for financial investors. This information can be used for many purposes among which risk management activities, trading strategies and option pricing. A common stylized fact of financial time series is that large absolute returns are more likely to be followed by large absolute returns. The same remark also applies for small returns. This stylized fact, known as volatility clustering, had motivated the work of Engle (1982) who proposes the ARCH(P) model where the conditional variance depends linearly on lagged squared innovations. In financial applications, to obtain a good fit, one needs a big integer for the parameter P and therefore more parameters to be estimated. In order to achieve a parsimonious parametrization, Bollerslev (1986) introduced the Generalized ARCH(P) model denoted by GARCH(P,Q). This model has the ability to sufficiently fit well asset prices even with small integers P and Q . During 1990s, another extension appears that integrates the fact that negative and positive return innovations impact differently future volatilities. Namely, the empirical evidence shows that impact of negative returns are more important than the positive ones (leverage effects). There are nowadays many GARCHs models that take into account this asymmetric effect, see Nelson (1991), Glosten, Jagannathan, and Runkle (1993), (GJR for short). Brownlees et al. (2011) provide also a good reference on this issue by make a forecasting comparative study.

A third extension is to consider many regimes of univariate GARCH models instead of one to generate enough skewness and kurtosis to match those of financial asset returns. The regimes may be independent or chosen through a Markov chain and the corresponding models are known as Markov regime switching GARCH models. Some papers related to this issue are Klaassen (2002), Marcucci (2005) and references therein. Switching GARCH models aim to capture the fact that volatility shocks are not persistent inside a regime (low or high volatility). Also, in these models, a small shock may be followed by a big shock and conversely a big shock may be followed by a small one. However, it is known that models with many parameters may have some problems of convergence in the estimation process or a lack of robustness in the out-of-sample evaluation because of the over-fitting phenomenon. For example, Marcucci (2005) demonstrates that Markov regime switching GARCH models do not dominate the single GARCH models with respect to VaR-based loss functions. He also finds the same results for volatility forecast accuracy for which no model clearly outperforms the others if short and long run time period are considered. Christoffersen and Jacobs (2004), for option pricing purpose, consider several single GARCH models. They find that the leverage GARCH model is not dominated with respect to other asymmetric models having more parameters in their formulation. All these results motivate us to provide a mathematical model that is based on a single Asymmetric GARCH framework. The first contribution of the paper is to propose a stochastic unit GARCH (SUGARCH hereafter) model defined by a standard GARCH model where some coefficients are multiplied by a

predictable stochastic factor having an expected value of one. In other words, the model may be seen as a GARCH model in "mean". In contrast to other asymmetric models, the SUGARCH(P,Q,O) class cannot include an integer which is greater than one for its third parameter O . If not, the GARCH model in "mean" property will be lost and the latter is the main difference with respect to other asymmetric models. Therefore, the class is always SUGARCH(P,Q,1) and may be denoted simply by SUGARCH(P,Q) where P, Q are strictly positive integers. Consequently, the model may be reliable for investors/market participants who want to use a model capturing both complexity (time-varying parameter) and simplicity (standard GARCH) for their investment or risk management purpose. The second contribution of the paper is empirical. It is found that asymmetric models perform well in the in-sample data if leverage effects are introduced on the GARCH parameter β . However, for the out-of-sample data, the relative performance is not too good and it is preferable to introduce leverage effects on the ARCH parameter α as usually done in the literature.

The remaining in this paper is organized around four sections. Section 2 presents the SUGARCH(P,Q) class and its competing asymmetric models. Section 3 describes the data and the methodology used in the empirical application. Section 4 presents the empirical results and the last section concludes.

2. Some AGARCH Models

2.1 The SUGARCH Class

The idea of SUGARCH class is to capture some properties of the standard GARCH model by taking into account the leverage effects. Specifically, I define it as a standard GARCH model with some coefficients multiplied by a predictable factor say v_t . The latter is such as $E(v_t)=1$. In the next step, to share some properties with GARCH models, some additional constraints are introduced on the conditional distribution of innovations which must belong to the class of symmetric distributions. This condition is not too restrictive since we may find some of them with more kurtosis than the normal distribution (i.e Student Law).

In this study, I am interested on forecasting conditional volatility on short horizon using daily data. So the conditional mean is supposed to be constant as in Klaassen (2002) or Marcucci (2005). Let S_t and r_t be respectively the security prices and security logarithmic returns, μ the conditional mean, then the SUGARCH(P,Q) class is defined by

$$r_t = \mu + \varepsilon_t = \mu + \sigma_t \eta_t, \quad (\varepsilon_t = \sigma_t \eta_t) \quad (1a)$$

$$\sigma_t^2 = \alpha_{0,t} + \sum_{i=1}^P \alpha_{i,t} \varepsilon_{t-i}^2 + \sum_{j=1}^Q \beta_{j,t} \sigma_{t-j}^2, \quad (1b)$$

$$\eta_t \sim IID f(0,1), \quad (1c)$$

where $\alpha_{0,t}$, $\alpha_{i,t}$, $\beta_{j,t}$ may be either constant ($\alpha_{0,t} = \alpha_0$, $\alpha_{i,t} = \alpha_i$, $\beta_{j,t} = \beta_j$) or stochastic. In the latter case, their expression is given by

$$\alpha_{0,t} = \alpha_0 v_t, \quad \alpha_{i,t} = \alpha_i v_t; \quad \beta_{j,t} = \beta_j v_t; \quad \text{with } v_t = 1 - \gamma \varepsilon_{t-1}. \quad (2)$$

The term $f(0,1)$ represents a (generic) symmetric density function with zero mean and unit variance and γ is the parameter that models asymmetry between negative and positive shocks. Note that, from Eq. (2), $E(v_t) = 1$ by using Eq. (1c) and a SUGARCH(P,Q) class has $2^{P+Q+1} - 1$ asymmetric models (Note 1). Hereafter, I will work only with the general SUGARCH model where each parameter is stochastic since the other special cases may be obtained from the same methodology. In this case, the conditional variance (1b) may be rewritten as

$$\sigma_t^2 = \left(\alpha_0 + \sum_{i=1}^P \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2 \right) - \gamma \varepsilon_{t-1} \left(\alpha_0 + \sum_{i=1}^P \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2 \right). \quad (3)$$

So to obtain an unconditional variance which is equal to the standard GARCH model, we need the expected value of the second term of Eq. (3) to be zero. For this, I assume the distribution of η_t to be symmetric with mean 0 which is a sufficient condition. In the appendix, it is then shown:

$$E(\sigma_t^2) = \bar{\sigma} = \frac{\alpha_0}{1 - \sum_{i=1}^P \alpha_i - \sum_{j=1}^Q \beta_j}. \quad (4)$$

Since the conditional volatility must be always positive and to obtain a covariance stationary model, the following constraints are made:

$$\alpha_0 > 0, \alpha_i \geq 0, i=1, \dots, P; \beta_j \geq 0, j=1, \dots, Q; \sum_{i=1}^P \alpha_i + \sum_{j=1}^Q \beta_j < 1 \quad (5a)$$

$$|\gamma| < \frac{1}{\max\{|\varepsilon_t|, t=1, \dots, N\}}, \varepsilon_t = r_t - \mu, \quad (5b)$$

where N is the number of observation returns.

Eq. (5a) gives the same GARCH constraints. Eq. (5b) handles the positivity of the stochastic factor $v_t = 1 - \gamma\varepsilon_{t-1}$ to ensure the positivity of the conditional variance, see (2). A general bound may be taken for

the innovations and then for the parameter γ . Here, I let the bound to depend on the data. The idea is to allow a big range of γ since the leverage effects are introduced by this parameter. In practice, since investors generally work with high frequency (intra or daily) observations, the conditional mean μ in Eq (1a) is often small and so the innovations may be approximated by asset returns in Eq. (5b) (Note 2). It is also expected that γ will be positive to integrate the fact that negative shocks impact more future volatilities than positive shocks.

On the other hand, when a big shock appears in the innovations, the volatility cannot persist for a long time. This is due to the multiplicative factor v_t that alternatively allows large and small volatility movements in a symmetric way since (ε_t) behaves as a fair game. Namely, we may have a small shock in a period of high volatility or a big shock in a period of low volatility. This feature is also shared by switching GARCH models that allow the volatility process to be in different levels. To see formally the link, note that if we have two regimes, as it is often the case in financial applications, the standard GARCH coefficients take two different values. Here, for more flexibility, the SUGARCH class allows stochastic coefficients v_t valued in $\mathbb{R}$ (Note 3)

Even if the coefficients are random, the framework is still similar to the standard GARCH model; which is an an interesting result. We have seen that both models share the same unconditional variance. The difference only appears *locally* where the stochastic factor v_t generates asymmetry effects and extreme movements for the conditional variance. This is the main feature that differentiates our model to other Asymmetric GARCH models which have also time-varying parameters but with an expected value different from 1. Consequently of these oscillations, the kurtosis of the distribution increases. Specifically, if the sixth moment of the asset return exists, it can be shown, see the appendix, that

$$E(\sigma_t^4) = \frac{\alpha_0^2 + E(\sigma_t^2) \left[2\alpha_0\alpha_1 + 2\alpha_0\beta_1 + \gamma^2\alpha_0^2 \right] + \gamma^2 E(\sigma_t^6) \left[15\alpha_1^2 + \beta_1^2 + 6\alpha_1\beta_1 \right]}{1 - 3\alpha_1^2 - \beta_1^2 - 2\beta_1\alpha_1 - 2\gamma^2(3\alpha_0\alpha_1 + \alpha_0\beta_1)} \quad (6)$$

For the forecast purpose, I use the simplest SUGARCH (1,1) class corresponding to $P=Q=1$. In this case, a closed formula for any multi-step-ahead volatility forecasts exists. Its form is similar to the standard GARCH(1,1) model where the difference appears only on the initial condition. Namely, we have for any integer $h \in \mathbb{N}^*$,

$$E_t(\sigma_{t+h}^2) = \frac{\alpha_0 [1 - (\alpha_1 + \beta_1)^{h-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{h-1} \sigma_{t+1}^2, \quad (7)$$

where σ_{t+1}^2 is the conditional volatility defined from (1b) i.e

$$\sigma_{t+1}^2 = \alpha_{0,t} + \alpha_{1,t} \varepsilon_t^2 + \beta_{1,t} \sigma_t^2$$

In this study, I only consider three of the seven asymmetric models of the SUGARCH(1,1) class given by the following conditional volatilities:

$$\sigma_{t+1}^2 = \alpha_0(1 - \gamma\varepsilon_t) + \alpha_1\varepsilon_t^2 + \beta_1\sigma_t^2 \quad (8a)$$

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1(1 - \gamma\varepsilon_t)\varepsilon_t^2 + \beta_1\sigma_t^2 \quad (8b)$$

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1\varepsilon_t^2 + \beta_1(1 - \gamma\varepsilon_t)\sigma_t^2 \quad (8c)$$

The four remaining asymmetric formulations use at least two stochastic parameters and I find that they do not produce any significant difference with respect to formulations (8a), (8b), (8c). Another reason for choosing these

three conditional volatilities is that they integrate leverage effects respectively on the constant α_0 , ARCH α_1 and GARCH β_1 parameter of the standard GARCH model. Therefore, it allows seeing which is the best way to capture asymmetry shocks on financial time series. The intuition of these three formulations is to model the negative correlation between past shocks (or returns) and future volatility. To the best of my knowledge, this problem is not studied in the literature and authors generally use a formulation similar to (8b) i.e asymmetry introduced through the ARCH parameter.

In the next section, I present two competing models (EGARCH, GJR), belonging to the class of asymmetric GARCH models. Additionally, I include the symmetric GARCH model which may give good results in the out-of-sample evaluation of volatility forecasts. Since the true conditional variance is not observable, square returns are used as a conditionally unbiased volatility proxy. The advantage of this proxy is that it ensures the correct ranking of predictive models in terms of the QLIKE loss functions, see Eq. (12). The other loss functions such that Mean Absolute Error, the Mean Square Errors on standard deviations may give some biases, see A. Patton (2011) for more information.

2.2 Competing Models

The EGARCH model: The Exponential GARCH (EGARCH) model is proposed by Nelson (1991). As its name indicates, the variable of interest is the logarithm of the conditional variance. It is defined as follows:

$$r_t = \mu + \varepsilon_t; \quad \varepsilon_t = \sigma_t \eta_t \quad (9a)$$

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^P g(\eta_{t-i}) + \sum_{j=1}^Q \beta_j \ln \sigma_{t-j}^2 \quad (9b)$$

$$g(\eta_{t-i}) = \alpha_i \left[|\eta_{t-i}| - E(|\eta_{t-i}|) \right] + \gamma_i \eta_{t-i}, \quad (9c)$$

where η_t was defined in Eq. (1c). The asymmetry between negative and positive shocks is modeled by the parameters (α_i, γ_i) , $i = 1, \dots, P$. Note that, if we are interested in forecasting, the conditional variance of EGARCH must depend on the distribution of η_t through $E(|\eta_{t-i}|)$. Since η_t has a generic symmetric density $f(0,1)$, I consider in this study two normalized distributions. The first is the Normal distribution:

$$f_{\eta_t}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad E(|\eta_t|) = \sqrt{\frac{2}{\pi}}.$$

The Student's t is the second distribution which has more kurtosis than the normal. Its density is given by

$$f_{\eta_t}(x, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)} \frac{1}{\sqrt{\pi(\nu-2)}} \left(1 + \frac{x^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad E(|\eta_t|) = \sqrt{\frac{\nu-2}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}.$$

At time t , the conditional expectation of Eq. (9b) is given for $P = Q = 1$ by

$$E_t(\ln \sigma_{t+1}^2) = \ln \sigma_{t+1}^2 = \alpha_0 + \gamma_1 \eta_t + \alpha_1 \left(|\eta_t| - E(|\eta_t|) \right) + \beta_1 \ln \sigma_t^2.$$

So, the one-step ahead volatility forecast is obtained by

$$\ln \hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\gamma}_1 \eta_t + \hat{\alpha}_1 \left(|\eta_t| - E(|\eta_t|) \right) + \hat{\beta}_1 \ln \sigma_t^2. \quad (10)$$

The multi-step-ahead volatility is derived recursively from $E_t(\ln \sigma_{t+h+1}^2) = \alpha_0 + \beta_1 E_t(\ln \sigma_{t+h}^2)$, due to Eq. (9c). Therefore, we have

$$\hat{\sigma}_{t+h+1}^2 = \exp \left[\hat{\alpha}_0 + \hat{\beta}_1 \ln \hat{\sigma}_{t+h}^2 \right],$$

where the initial condition is given by (10) and all parameters are estimated by the Maximum Likelihood method.

The GJR GARCH model: It is proposed by Glosten, Jagannathan and Runkle (1993), for short GJR model. The conditional variance of the asset return is defined by

$$r_t = \mu + \varepsilon_t; \quad \varepsilon_t = \sigma_t \eta_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^P \alpha_i \varepsilon_{t-i}^2 + \sum_{k=1}^O \gamma_k \varepsilon_{t-k}^2 I_{\{\varepsilon_{t-k} < 0\}} + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2$$

where $I_{\{\varepsilon_{t-k} < 0\}}$ is an indicator function that takes the value 1 if $\varepsilon_{t-k} < 0$ and 0 otherwise. Some constraints need to be made to insure strict positivity of the volatility. For the simplest model GJR GARCH (1,1,1), we have $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\alpha_1 + \gamma_1 \geq 0$.

Since η_t is symmetric, the multi-step-ahead volatility forecast is recursively obtained from

$$E_t(\sigma_{t+h+1}^2) = \alpha_0 + (\alpha_1 + \frac{\gamma_1}{2} + \beta_1)E_t(\sigma_{t+h}^2).$$

Therefore, we have at time t

$$\hat{\sigma}_{t+h+1}^2 = \hat{\alpha}_0 + (\hat{\alpha}_1 + \frac{\hat{\gamma}_1}{2} + \hat{\beta}_1)\hat{\sigma}_{t+h}^2; \quad (11a)$$

$$\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \varepsilon_t^2 + \hat{\gamma}_1 \varepsilon_t^2 I_{\{\varepsilon_t < 0\}} + \hat{\beta}_1 \sigma_t^2. \quad (11b)$$

The GARCH model: The model is proposed by Bollerslev (1986) as an extension of the ARCH model of Engle (1982). It is a special case of SUGARCH (set for any t $\alpha_{0,t} = \alpha_0$, $\alpha_{i,t} = \alpha_i$, $\beta_{j,t} = \beta_j$) and GJR-GARCH models. Its unconditional variance is given by Eq. (4) and the multi-step-ahead formula may be obtained from Eqs. (7) and (8a) by setting $\gamma = 0$.

Note that the literature of asymmetric GARCH models allows more than one parameter for the leverage effects without significantly affecting the structure of the model. This is not the case of SUGARCH class where the parameter O is restricted to be always one. The reason is that if the stochastic factor v_t contains more parameters ε_{t-2} , ε_{t-3} , ..., the explicit formula for the unconditional variance is lost for some formulations. In this case, the second term of Eq. (3) may involve expressions such that $E(\varepsilon_{t-i} \varepsilon_{t-j}^2)$, $E(\varepsilon_{t-i} \sigma_{t-j}^2)$ with $j < i$ and those terms will be different from zero. Therefore, the acronym SUGARCH will be lost. On the other hand, if a symmetric distribution for η_t is considered as a strong requirement for financial returns, we can avoid the distributional assumptions and estimate the model by using Quasi Maximum Likelihood method.

3. Data and Methodology

I consider the S&P 500 daily time series, adjusted for dividends, to evaluate the performance of the different models presented above. The sample period is from January 2, 2002 to December 31, 2010 corresponding to $N=2267$ daily observations. Table 1 gives the descriptive statistics of the index with respect to its asset returns $r_t = \ln S_t - \ln S_{t-1}$ where S_t represents the spot price at time t .

Table 1. Descriptive statistics of S & P 500 returns

Mean	Std. deviation	Min	Max	Skewness
0.0038	1.3791	-9.469	10.957	-0.139
Kurtosis	JB-stat	LB(20)-stat	LM(10)-stat	
11.854	7.410 (< 0.001)	97.171 (4.0210^{-12})	666 (0)	

JB-stat represents the Jarque-Bera statistic for normality test. LB(20)-stat corresponds to the residual autocorrelation test of Ljung-Box including up to 20 lags. LM(10)-stat is a statistic which examines for the presence of ARCH effects where 10 lags are used for the squared sample residuals. I give in brackets the corresponding p-values of the different statistics.

The table shows that the mean return of the S&P 500 index is positive and small. The standard deviation is also small (1.38%). The maximum (minimum) return is given by 10.957 % (-9.469 %). Extreme movements appear more frequently since the null hypothesis of normal distribution for the unconditional return is highly rejected even at the 1% significance level. The same thing appears for the null hypotheses of no serial correlation as well as the null of no ARCH effects. Both hypotheses are rejected with p-values close to zero. These results from Table 1

suggest to use GARCH models to take into account the excess of kurtosis and the presence of heteroskedasticity. In the previous section, I reviewed some of them. To make forecast, I estimate the parameters of each model by using the Maximum Likelihood method where the conditional distribution for innovations is either normal or Student-t. Then, the future volatilities are forecast and some quantiles (Value-at-Risk) are also determined. The time horizon h belongs to $\{1,2,5,10\}$, corresponding to 1 day, 2 days, one week and two weeks, respectively. The literature usually compares the relative performance of volatility models around a statistical loss function or an economic loss function. Only the former is considered with the Quasi-Likelihood loss function, as a criterion, defined as follows

$$QLIKE = \frac{1}{N-h-T+1} \sum_{t=T}^{N-h} \left(\log \hat{\sigma}_{t+h|t}^2 + \frac{\varepsilon_{t+h}^2}{\hat{\sigma}_{t+h|t}^2} \right), \quad (12)$$

where T and N are length of the in-sample data and total sample, respectively.

QLIKE function shares robustness on ranking the models with respect to an unbiased estimator of the unknown conditional variance, see for example Patton (2011).

The in-sample data spans the period January 2, 2002 to October 2, 2008 corresponding to $T = 1699$ ($0.75 \cdot N$) and the remaining (25% of the data) is used as the out-of-sample data. The parameter h gives the horizon forecast used to compare models. In this study, I focus only on the QLIKE loss metric for some reasons given by Brownlees et al. (2011). The authors note that QLIKE may be rewritten without loss of generality by

$$QLIKE = \frac{1}{N-h-T+1} \sum_{t=T}^{N-h} \left(-\log \frac{\varepsilon_{t+h}^2}{\hat{\sigma}_{t+h|t}^2} + \frac{\varepsilon_{t+h}^2}{\hat{\sigma}_{t+h|t}^2} \right)$$

and so is a combination of i.i.d terms (η_t). Another reason is that QLIKE penalizes small volatility forecast (close to zero).

Even if metric criteria are important, it is useful to have statistical tests that assess if the difference between loss functions of two competing models is significant or not. For this, I consider the test of equal predictive ability (EPA) of Diebold and Mariano (1995) (DM hereafter). Let d_t be the loss differential between the two competing forecasts i.e:

$$d_t = g(\sigma_t, \hat{\sigma}_{t|t-h}^i) - g(\sigma_t, \hat{\sigma}_{t|t-h}^j),$$

where g is given by (12). The above terms $\hat{\sigma}_{t|t-h}^i, \hat{\sigma}_{t|t-h}^j$ represent the multi-step-ahead volatility forecasts (horizon h) of σ_t from model i and j , made at time $t-h$. The asymptotic DM test is based on the mean of loss differential series (d_t), $t = T+h, \dots, N$. If the latter is covariance stationary and short memory, the authors show that

$$\sqrt{n_h}(\bar{d} - \mu) \xrightarrow{d} \mathbf{N}(0, V(\bar{d})), \quad \bar{d} = \frac{1}{n_h} \sum_{t=T+h}^N d_t, \quad n_h = N-h-T+1,$$

where μ represents the population mean loss differential. The variance is estimated by $\hat{V}(\bar{d}) = \frac{1}{n_h} (\hat{\gamma}_0 + 2 \sum_{k=1}^q \omega_k \hat{\gamma}_k)$. Here, $\hat{\gamma}_k$ represents an estimate of the k -th order auto-covariance of the series (d_t), q the truncation lag and ω_k the lag windows. I follow Marcucci (2005) by taking $q = 4 * (\frac{N}{100})^{2/9}$ and $w_k = 1 - \frac{k}{q+1}$. So the DM test of equal predictive accuracy ($\mu = 0$) is given asymptotically by

$$\frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})/n_h}} \sim \mathbf{N}(0,1). \quad (13)$$

After assessing the loss functions of competing models and analyzing their statistical significance, I evaluate volatility forecasting performance in a financial risk management setting. For this, I calculate the Value-at-Risk which is the money-loss in a portfolio that is expected to occur over a pre-determined horizon (h) and with a pre-determined degree of confidence (α). It may be seen also as a quantile of the portfolio (conditional) distribution. Precisely, consider equation (1a) and let $F_{t,h} = P(r_t \leq r | \mathbf{F}_{t-h})$, $G_{t,h} = P(\eta_t \leq \eta | \mathbf{F}_{t-h})$ be

the cumulative distribution of r_t and η_t , respectively, given the information set $\mathbf{F}_{t-h}$. The VaR with a tail probability $\alpha \in]0, 1[$ and horizon h , denoted $VaR_t(\alpha, h)$, is calculated at time $t-h$, by

$$VaR_t(\alpha, h) = F_{t,h}^{-1}(\alpha) = \mu + \sigma_t G_{t,h}^{-1} = \mu + \sigma_t G^{-1}(\alpha).$$

The last equality is explained by the fact that (η_t) is i.i.d, see (1c) and so $G_{t,h}^{-1} = G^{-1}$. The difference between models appears on how σ_t is forecast from $t-h$. After getting estimates of $VaR_t(\alpha, h)$, $t = T+h, \dots, N$ for each model, I define the following two loss functions for investors with long positions

$$\hat{\alpha} = \frac{1}{N-T-h+1} \sum_{t=T+h}^N 1_{(r_t < VaR_t(\alpha, h))} \quad (14a)$$

$$\hat{Q}(\alpha) = \frac{1}{N-h-T+1} \sum_{t=T+h}^N \left(\alpha - 1_{\{r_t < VaR_t(\alpha, h)\}} \right) (r_t - VaR_t(\alpha, h)). \quad (14b)$$

The best model with respect to the loss function (14a) is the one that minimizes the function $d(\hat{\alpha}) = |\hat{\alpha} - \alpha|$ even if it is preferable to have low values $\hat{\alpha} \leq \alpha$. I refer to Christoffersen (1998) for some statistical tests based on the coverage probability. The second loss function was proposed by Koenker and Bassett (1978), hereafter KB. It penalizes more heavily the observations for which there is a violation of VaR constraints. I also evaluate the performance of the competing models with respect to investors having short positions (Note 4).

4. Empirical Results

The results are based on the S&P 500 daily data adjusted for dividends. The data is extracted from yahoo finance web site (Note 5). I recall that all parameters are estimated by the Maximum Likelihood method with a Gaussian and a Student's t distribution for η_t and also the in-sample data ranges from January 2, 2002 to October 2, 2008 corresponding to 1699 daily observations. The labels ASUG, BSUG and CSUG in the Table 2 correspond to the conditional variance given by Eqs. (8a), (8b), (8c) respectively.

As expected, the estimated value of γ (leverage parameter) is such that negative shocks impact more future volatilities than positive ones. Accordingly, it is positive for SUGARCH and GJR models and negative for the EGARCH model. Also, the persistence of shocks on volatility ($\alpha_1 + \beta_1$) is high (> 0.990) for all models as it is usual in the financial time series. It is noted that introducing leverage effects on the constant of GARCH model (ASUG) or on the ARCH parameter (BSUG) does not give significant difference on the estimated parameters with respect to a standard GARCH model. However, if the asymmetry between negative and positive shocks is modeled through the GARCH parameter (CSUG model), the difference becomes clear since $\hat{\beta}_1 > 0.96$, (see Figure 1 for illustration). Figure 1 shows also the necessity to have time varying parameters for the standard GARCH model since its parameters have big oscillations through the time evolution. The use of SUGARCH class solves this problem due to its stochastic parameters.

Table 2. Estimated parameters of the different models

Models	mu	Alpha0	Alpha1	Beta1	Leverage	Dof
ASUG_N	0.033 (0.021)	0.008 (0.005)	0.062 (0.011)	0.931 (0.014)	0.097 (0.014)	—
ASUG_T	0.044 (0.022)	0.005 (0.003)	0.061 (0.011)	0.935 (0.013)	0.097 (0.006)	10.09 (0.01)
BSUG_N	0.028 (0.029)	0.008 (0.005)	0.060 (0.009)	0.933 (0.012)	0.097 (0.11)	—
BSUG_T	0.041 (0.03)	0.005 (0.005)	0.060 (0.024)	0.936 (0.026)	0.097 (0.74)	10.69 (0.71)
CSUG_N	0.022 (0.023)	0.005 (0.003)	0.035 (0.009)	0.962 (0.010)	0.089 (0.010)	—
CSUG_T	0.044 (0.017)	0.003 (0.002)	0.031 (0.009)	0.966 (0.010)	0.092 (0.011)	11.47 (0.010)
EGARCH_N	0.007 (0.020)	0.0003 (0.003)	0.078 (0.015)	0.986 (0.002)	-0.113 (0.012)	—
EGARCH_T	0.017 (0.019)	-0.003 (0.003)	0.071 (0.017)	0.990 (0.003)	-0.111 (0.013)	13.461 (2.409)
GJR_N	0.002 (0.020)	0.009 (0.002)	0.000 (0.011)	0.938 (0.011)	0.107 (0.017)	—
GJR_T	0.018 (0.019)	0.006 (0.002)	0.000 (0.014)	0.941 (0.012)	0.108 (0.020)	11.716 (2.236)
GARCH_N	0.034 (0.020)	0.008 (0.002)	0.063 (0.009)	0.932 (0.010)	—	—
GARCH_T	0.044 (0.019)	0.004 (0.003)	0.063 (0.011)	0.937 (0.011)	—	9.623 (1.753)

Note: Each model has been estimated by Maximum likelihood method with a Normal (_N) and a Student's t(_T) distribution for η_t . Asymptotic standard errors are in parentheses. Dof stands for degree of freedom.

Next, I give in-sample evaluation of the different models with respect to the QLIKE loss function as well as for the log likelihood function (LLF), $AIC = -2\log(LLF)/T + 2k/T$ and $BIC = -2\log(LLF)/T + (k/T)\log(T)$ criteria where T represents the in-sample size and k the number of model's parameters. The results are shown in Table 3.

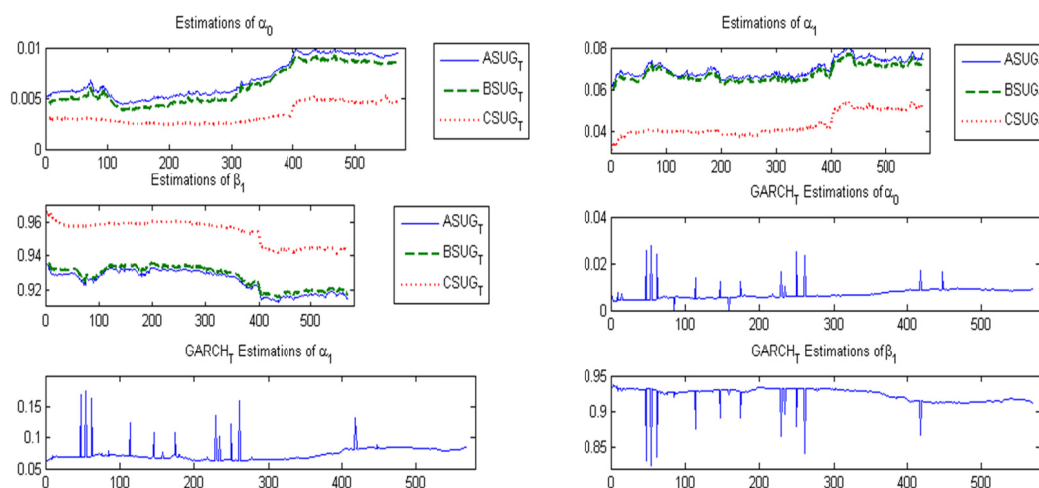


Figure 1. Comparison of recursive estimated parameters for some models ($SUGARH_T$ vs $GARCH_T$). The number of observations is 568 beginning from $T=1699$ to $N=2266$.

Table 3. In-sample diagnostic for the different models

Models	LLF	AIC	BIC	QLIKE
ASUG_N	-2286.88	2.698	2.714	24656043.71
ASUG_T	-2267.54	2.676	2.695	2469815.15
BSUG_N	-2277.87	2.687	2.703	2434990.48
BSUG_T	-2259.692	2.667	2.686	2440255.29
CSUG_N	-2256.51	2.662	2.678	2362420.94
CSUG_T	-2238.72	2.642	2.661	2368878.219
EGARCH_N	-2258.825	2.665	2.681	2370270.553
EGARCH_T	-2240.275	2.644	2.663	2375141.984
GJR_N	-2256.050	2.662	2.678	2360839.063
GJR_T	-2240.896	2.645	2.664	2367111.968
GARCH_N	-2287.273	2.697	2.710	2466936.088
GARCH_T	-2267.389	2.675	2.691	2473089.834

Note: This table presents the Log likelihood function(LLF), the Akaike information criterion(AIC) and the Schwarz criterion (BIC). QLIKE is the loss function defined in Eq. (12). Numbers in boldface indicate the best values.

The student's t distribution for η_t gives better fitting than normal distributions. The three best models are given respectively by CSUG_T, EGARCH_T, GJR_T for LLF criterion. The ranking order is also the same for AIC and BIC criteria. This point highlights the finding that normalized financial returns $\eta_t = (r_t - \mu) / \sigma_t$ still have a heavy tail distribution, however with less kurtosis than the unconditional return $(r_t - \mu)$ distribution. If QLIKE is considered as a measure, the normal distribution for η_t is better than the Student distribution. The three best models are given respectively by GJR_N, CSUG_N, GJR_T. Overall, I note that introducing asymmetry through β_1 (CSUG model) instead of α_0 or α_1 (ASUG, BSUG model) improves accuracy in the in-sample fitting. Since the model structure has similarities with other (A) GARCH models in the sense that current volatility depends on past volatilities (GARCH parameters) and past innovations (ARCH parameters), we may expect that the same technique to work also for those models to better approximate financial data. The next step is to see the out-of-sample performance of the different models, part that interests more investors and market participants.

For this purpose, I compare the relative performance of the different volatility models in three ways. The first uses a metric loss function, the second is based on directional accuracy tests whereas the third focuses on risk management purpose. The out-of-sample data ranges from October 3, 2008 to December 31, 2010 and represents twenty six months of data (567 observations).

Table 4 shows the comparison between competing models for the QLIKE loss function (metric criterion) for all horizons $h = 1, 2, 5, 10$. The true m-step-ahead variance is approximated by the squared return of the forecast horizon that means $(\ln S_{t+h} - \ln S_t)^2$.

It is seen that no model outperforms the others. For example, for daily volatility forecasts ($h = 1$), GJR_T model gives the best results while for the remaining step-ahead volatility forecasts ($h = 2, 5, 10$), the best model is BSUG_T. Another point is that the assumption of normal distribution for innovations generates satisfactory results specifically for daily volatility forecasts. However, for longer horizon, minimal values for the QLIKE loss function are obtained with a Student's t distribution. Also, a good in-sample performance does not imply a good out-of-sample performance. We see previously that CSUG_T was the best model from in-sample performance while it gives here no satisfactory results on the out-of-sample evaluation when multi-step-ahead volatility forecasts are considered. Overall, the best model is now BSUG_T.

Table 4. Out-of-sample evaluation of volatility forecasts for $h=1, 2, 5, 10$

Models	H	QLIKE	h	QLIKE	h	QLIKE	h	QLIKE
ASUG_N	1	557978.6	2	835203.8	5	1627324.9	10	2911320.9
ASUG_T	1	558362.5	2	831520.9	5	1606663.9	10	2863105.6
BSUG_N	1	553845.2	2	826436.8	5	1599190	10	2878972.9
BSUG_T	1	554372.2	2	820814.5	5	1569805.8	10	2807348.9
CSUG_N	1	547919.9	2	863287.7	5	1744494.3	10	3171634.2
CSUG_T	1	549464	2	865474.9	5	1746438	10	3185808.7
EGARCH_N	1	560398.8	2	836046.7	5	1616423.3	10	2887507.8
EGARCH_T	1	563255.8	2	843229.9	5	1675784.8	10	3124184.1
GJR_N	1	544429.1	2	837609	5	1648441.4	10	3034470.1
GJR_T	1	543582.2	2	824712.4	5	1607949.9	10	2924896.5
GARCH_N	1	556372.8	2	835519.6	5	1641501.4	10	2951462
GARCH_T	1	557947.8	2	829282.1	5	1594605.2	10	2821015

Note: This table presents the Quasi likelihood (QLIKE) loss function, see, (12). Numbers in boldface give the minimal (best) value for each group.

The Diebold and Mariano test, see Eq. (13) is now adopted to further examine the statistical significance from two competing models i and j . The findings from the DM-test statistics across all models and forecast horizons are available. Table 5 presents the results obtained from *BSUG_T* and *GJR_T* taken as benchmarks where forecast horizons are given respectively by $h = 5, 10$ (Note 6).

Table 5. Diebold-Mariano Test with *BSUG_T* and *GJR_T* as benchmarks

Bench. <i>BSUG_T</i>	horizon	QLIKE	Bench. <i>GJR_T</i>	horizon	QLIKE
ASUG_N	5	-2.87**	ASUG_N	1	-2.37*
P-values		(0.004)	P-values		(0.018)
ASUG_T	5	-1.73	ASUG_T	1	-2.54*
P-values		(0.08)	P-values		(0.011)
BSUG_N	5	-5.50**	BSUG_N	1	-2.30*
P-values		(0.00)	P-values		(0.021)
CSUG_N	5	-4.85**	BSUG_T	1	-2.59**
P-values		(0.00)	P-values		(0.009)
CSUG_T	5	-4.40**	CSUG_N	1	-0.689
P-values		(0.00)	P-values		(0.49)
EGARCH_N	5	-0.67	CSUG_T	1	-0.843
P-values		(0.49)	P-values		(0.40)
EGARCH_T	5	-1.21	EGARCH_N	1	-1.84
P-values		(0.22)	P-values		(0.06)
GJR_N	5	-4.78**	EGARCH_T	1	-1.57
P-values		(0.00)	P-values		(0.11)
GJR_T	5	-2.72**	GJR_N	1	-0.13
P-values		(0.006)	P-values		(0.84)
GARCH_N	5	-3.36**	GJR_T	1	-2.09*
P-values		(0.00)	P-values		(0.03)
GARCH_T	5	-0.16	GARCH_N	1	-2.63**
P-values		(0.86)	P-values		(0.008)

Note: Bench. stands for Benchmark. * and ** represent the DM statistics for which one can reject the null hypothesis of equal predictive accuracy at 5% and 1%, respectively.

As expected, the Diebold and Mariano (DM) test confirms results obtained from the previous table. For the benchmark model *GJR_T*, it is seen that all DM statistic values are negative showing that its loss function is the smallest for $h = 1$. On the other hand, the table shows that the null hypothesis of equal predictive ability is rejected for the following competing models ASUG, BSUG, GARCH. For the remaining, the difference between loss functions is not significant at the 5% level. When *BSUG_T* is now the benchmark, similar results are

obtained. It performs well in terms of volatility forecasts with respect to other models for the horizons $h = 2, 5, 10$. For example, for $h = 5$, the DM test is negative and the test is statistically significant at 1% level when a comparison is made with respect to *GJR*, *GARCH_N*, *ASUG_N*, *BSUG_N*, *CSUG* models. I have also the same conclusion for the other horizons $h = 2, 10$. Finally, it is noticed that all the best models for a given horizon do not give significant difference with respect to EGARCH model. So even if the latter does not outperform the others, it gives satisfactory results.

Finally, I compare the performance of the different models with respect to the two loss functions defined in (14a), (14b). The coverage rate of the VaR is $\alpha = 0.01$ and the distribution of η_t is assumed to be either Normal or a Student. A general finding is that all the models have problems to give good realized VaR forecasts when the horizon step is a week or two weeks. If the horizon is however 1 or 2 days, results are satisfactory. For the risk management purpose, I only analyze one side of the conditional return distribution since if an investor takes a long (short) position, only extreme negative (positive) returns would matter for him.

Table 6. Out-of-sample Evaluation: 99 % VaR, Long Position

Models	horizon	PF	KB	horizon	PF	KB
ASUG_N	1	2.991E-02	4.6927E-02	2	5.300E-02	9.359E-02
ASUG_T	1	5.2910E-03	4.9117E-02	2	3.003E-02	6.854E-02
BSUG_N	1	2.64E-02	4.670E-02	2	4.77E-02	8.860E-02
BSUG_T	1	3.5273E-03	4.927E-02	2	2.826 E-02	6.694E-02
CSUG_N	1	3.5273E-02	5.8116E-02	2	6.360E-02	1.165E-01
CSUG_T	1	1.23409E-02	5.24609E-02	2	4.5930E-02	9.255E-02
EGARCH_N	1	3.1746E-02	5.810E-02	2	6.183E-02	9.99E-02
EGARCH_T	1	7.0547E-03	5.8360E-02	2	2.120E-02	7.281E-02
GJR_N	1	2.2465E-02	4.6826E-02	2	5.477E-02	9.02E-02
GJR_T	1	7.054 3E-03	4.860E-02	2	3.180E-02	6.76E-02
GARCH_N	1	2.6452E-02	4.671E-02	2	5.477E-02	9.306E-02
GARCH_T	1	5.2910E-03	4.971E-02	2	3.00E-02	6.695E-02

Note: This table presents the percentage proportion of failures (PF) and Koenker and Bassett (KB) loss function for the 99 % VaR failure processes at one and two-step-ahead. Numbers in boldface show the best values.

Table 6 shows that for long position and $h=1$, CSUG_T gives the best estimation of the theoretical value $\alpha = 1\%$. It is followed by GJR_T and EGARCH_T models. The latter becomes the best model for the two day-horizons followed by the BSUG_T model. Since, the PF loss function does not take into account the magnitude of VaR violations (same weight), I add Koenker and Basset (KB) loss function to remedy this disadvantage. In this case, VaR violations (no VaR violations) are weighted by $1 - \alpha$ (α). So the Koenker and Basset (KB) loss function penalizes heavily prediction with VaR violations since α is equal in practice to 1% or 5%. It is found that the outperforming model is now BSUG_N followed by GARCH_N for $h=1$ as well as for $h=2$. It is also noted that for $h=2$, all models have more VaR violations than the theoretical value 1%. In other words, it is preferable to work only on daily observations for the long position.

Table 7. Out-of-sample Evaluation: 99 % VaR, Short Position

Models	horizon	PF	KB	horizon	PF	KB
ASUG_N	1	1.05E-02	4.77E-02	2	3.00E-02	5.97E-02
ASUG_T	1	3.52E-03	4.96E-02	2	7.06E-03	5.10E-02
BSUG_N	1	1.05E-02	4.44E-02	2	2.47E-02	5.29E-02
BSUG_T	1	0	4.88E-02	2	7.06E-03	5.13E-02
CSUG_N	1	5.29E-03	4.03E-02	2	2.65E-02	4.59E-02
CSUG_T	1	3.52E-03	4.35E-02	2	7.06E-03	4.46E-02
EGARCH_N	1	1.41E-02	4.60E-02	2	2.65E-02	5.52E-02
EGARCH_T	1	3.52E-03	5.41E-02	2	5.30E-03	5.53E-02
GJR_N	1	1.41E-02	4.37E-02	2	2.47E-02	4.97E-02
GJR_T	1	3.53E-03	4.86E-02	2	5.30E-03	4.86E-02
GARCH_N	1	1.05E-02	4.79E-02	2	3.18E-02	5.95E-02
GARCH_T	1	1.76E-03	4.99E-02	2	7.06E-03	5.10E-02

Note: This table presents the percentage proportion of failures (PF) and Koenker and Bassett (KB) loss function for the 99 % VaR failure processes at one and two-step-ahead. Numbers in boldface give the best value.

Table 7 shows results for a short position. For daily VaR predictions, there are three best models for the PF criterion given by ASUG_N, BSUG_N and GARCH_N models. For the same horizon, the best accuracy for KB loss function is obtained from CSUG_N. For the horizon $h = 2$, CSUG_T is the model that minimizes also the loss function KB. For the PF loss criterion, the best models are a group given by ASUG_T, BSUG_T, CSUG_T and GARCH_T model. Overall, BSUG and CSUG models give satisfactory results with respect to VaR-based loss functions.

5. Conclusion

This paper has sought to re-examine the volatility forecasting literature by improving the standard GARCH model. The latter is extended by introducing asymmetry between negative and positive shocks. This extension, in contrast to other AGARCH models, does not change significantly the structure of the standard GARCH models. Also, I analyze what is the good way to capture leverage effects in financial time series. Our findings are summarized as follows. For in-sample fitting, the best model comes from SUGARCH class and it is the one obtained by modifying β_1 instead of α_1 parameter for asymmetric effects. Consequently, the GARCH parameter is more flexible than the ARCH parameter and is more suitable for financial asset prices. CSUG model is also the one inside the SUGARCH class that gives estimates significantly different from GARCH model.

For the out-of-sample evaluation, good results usually come from the SUGARCH class. For example, for the loss function QLIKE as a criterion, it is found that GJR is the best model for daily horizon but for $h = 2$, $h = 5$ and $h = 10$, it is better to work with the BSUG model.

These findings are also confirmed by the second criterion which is the statistical test defined by Diebold and Mariano. The latter test additionally shows that, all best models, in terms of loss ranking, do not give significant difference with respect to the EGARCH model.

Finally, I investigate performance of the different models with respect to loss functions based on Value at Risk predictions. The first criterion is based on the coverage probability or the number of VaR violations. The obtained results depend on the investor's position. If long positions are considered, CSUG_T and EGARCH_T models gives respectively the best results for $h = 1$ and $h = 2$. However, for short position it is preferable to work with the SUGARCH class (ASUG, BSUG, CSUG). The second measure (KB loss function) integrates both the number and size of VaR violations. So it is more relevant than the coverage probability. For this criterion, the best models belong to the SUGARCH class (CSUG or BSUG model) independently on the investor position (long or short). These good results from SUGARCH class may be explained by inheritance of the standard GARCH model since they almost share the same formula for forecasting volatility where the difference only appears on the initial condition that integrates asymmetry. Another explanation is to see that SUGARCH class has some similarities with Mixture or Markov regime switching GARCH models and the literature has demonstrated that those models may give interesting results especially when economic changes appear on the interval of study as in our case (subprime crisis).

References

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327. [http://dx.doi.org/10.1016/0304-4076\(86\)90063-1](http://dx.doi.org/10.1016/0304-4076(86)90063-1)
- Brownlees, C., Engle, R., & Kelly, B. (2011). A Practical Guide to Volatility Forecasting Through Calm and Storm. *Journal of Risk*, 14(2), 1-20
- Christoffersen P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841-862. <http://dx.doi.org/10.2307/2527341>
- Christoffersen, P., & Jacobs, K. (2004). Which garch model for option valuation? *Management Science*, 50, 1204-1221. <http://dx.doi.org/10.1287/mnsc.1040.0276>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253-263.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, 50, 987-1008. <http://dx.doi.org/10.2307/1912773>
- Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 347-370. <http://dx.doi.org/10.1111/j.1540-6261.1993.tb05128.x>

- Klaassen, F. (2002). Improving garch volatility forecasts. *Empirical Economics*, 27, 363-394. <http://dx.doi.org/10.1007/s001810100100>
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50. <http://dx.doi.org/10.2307/1913643>
- Marcucci, J. (2005). Forecasting stock market volatility with regime-switching garch models. *Studies in Nonlinear Dynamics and Econometrics*, 9(4), 1-55. <http://dx.doi.org/10.2202/1558-3708.1145>
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach.. *Econometrica*, 59, 347-370. <http://dx.doi.org/10.2307/2938260>
- Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246-256. <http://dx.doi.org/10.1016/j.jeconom.2010.03.034>

Notes

Note 1. Since each parameter may take two forms, the cardinal is 2^{P+Q+1} . The minus 1 corresponds to the standard GARCH model where there is no stochastic parameter and so no asymmetry.

Note 2. In the implementation, I use $|\gamma| < \frac{0.9}{\max\{|r_t|, t=1, \dots, T\}}$.

Note 3. I also tried other formulations of SUGARCH models that are close to regime switching models. Namely, I define v_t by $v_t = 1 - \gamma * \text{sign}(\varepsilon_{t-1})$, so giving two values for v_t . The corresponding models give also the same unconditional variance than the standard GARCH model and have closed formula for the kurtosis.

Note 4. The loss function becomes in this case $\frac{1}{N-h-T+1} \sum_{t=T+h}^N \left(1_{\{r_t > VaR_t(1-\alpha, h)\}} - \alpha \right) \left(r_t - VaR_t(1-\alpha, h) \right)$

Note 5. For robustness, the same treatment is also made for CAC 40 index with similar conclusions.

Note 6. Due to space constraints, all results have not been included. The other ones can be downloaded from <http://sites.google.com/site/makonte/>

Appendix

Unconditional variance of SUGARCH (P,Q)

I have due to the hypothesis $\eta_t \sim f(0,1)$ and the law of iterated expectations

$$E(\varepsilon_{t-1}) = E[(\varepsilon_{t-1} \cdot \varepsilon_{t-i}^2)] = E[(\varepsilon_{t-1} \cdot \sigma_{t-j}^2)] = 0, \quad \forall i \geq 2, j \geq 1.$$

Since the η_t distribution is supposed to be symmetric, I also have $E[(\varepsilon_{t-1})^{2n+1}] = 0 \quad \forall n \in \mathbb{N}$. These equations imply then

$$E\left[\gamma \varepsilon_{t-1} \left(\alpha_0 + \sum_{i=1}^P \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2 \right)\right] = 0.$$

So, equation (3) gives

$$E(\sigma_t^2) = \alpha_0 + \sum_{i=1}^P \alpha_i E(\varepsilon_{t-i}^2) + \sum_{j=1}^Q \beta_j E(\sigma_{t-j}^2),$$

and equation (4) is obtained by using the fact that $E(\sigma_{t-j}^2) = E(\varepsilon_{t-i}^2)$ for any integers i, j .

Kurtosis of SUGARCH(1,1) class

For $P=1, Q=1$, I consider the model of SUGARCH(1,1) class defined by the following conditional volatility

$$\sigma_t^2 = \alpha_0 v_t + \alpha_1 v_t \varepsilon_{t-1}^2 + \beta_1 v_t \sigma_{t-1}^2 = \left(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \right) v_t, \quad v_t = 1 - \gamma \varepsilon_{t-1}.$$

$$\begin{aligned}
E(\sigma_t^4) &= E\left[(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2 (1 - \gamma \varepsilon_{t-1})^2\right] \\
&= E\left[(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2 (1 - 2\gamma \varepsilon_{t-1} + \gamma^2 \varepsilon_{t-1}^2)\right] \\
&= E\left[(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2 (1 + \gamma^2 \varepsilon_{t-1}^2)\right].
\end{aligned}$$

The last equation is obtained by the fact that $E\left[\varepsilon_{t-1}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2\right] = 0$ which comes from

$$E(\varepsilon_{t-1}) = E(\varepsilon_{t-1}^3) = E(\varepsilon_{t-1}^5) = E(\varepsilon_{t-1} \sigma_{t-1}^2) = E(\varepsilon_{t-1} \sigma_{t-1}^4) = E(\varepsilon_{t-1}^3 \sigma_{t-1}^2) = 0.$$

Therefore,

$$E(\sigma_t^4) = E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2 + \gamma^2 E\left[\varepsilon_{t-1}^2 (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^2\right].$$

It remains to develop the last equality and then to use the following points coming from $\eta_t \sim N(0,1)$

$$\begin{aligned}
E(\varepsilon_{t-1}^2) &= E(\sigma_{t-1}^2), & E(\varepsilon_{t-1}^4) &= 3E(\sigma_{t-1}^4), & E(\varepsilon_{t-1}^2 \sigma_{t-1}^2) &= E(\sigma_{t-1}^4) \\
E(\varepsilon_{t-1}^6) &= 15E(\sigma_{t-1}^6), & E(\varepsilon_{t-1}^2 \sigma_{t-1}^4) &= E(\sigma_{t-1}^6), & E(\varepsilon_{t-1}^4 \sigma_{t-1}^2) &= 3E(\sigma_{t-1}^6),
\end{aligned}$$

To obtain (6). The kurtosis is then deduced by the following formula

$$\kappa_{SUGARCH} = \frac{E(r_t^4)}{E(r_t^2)^2} = \frac{3E(\sigma_t^4)}{E(\sigma_t^2)^2}.$$

The expression of the kurtosis is not explicit. The unconditional variance $E(\sigma_t^2)$ is known, see (4), but not the term $E(\sigma_t^4)$, see (6), which depends on the unknown moment $E(\sigma_t^6)$. Anyway, for $\gamma \neq 0$, I can assert that the SUGARCH kurtosis is bigger than the standard GARCH(1,1) model corresponding to $\gamma = 0$ since they share the same denominator while $E(\sigma_t^4)$ of Eq. (6) is an increasing function of γ^2 .