

QML Estimation of GARCH(1,1) Process

Mona Samy Elkhoully*

*Assistant Lecturer at Statistics, Mathematics and Insurance Department
Faculty of Commerce –Port Said University

ملخص البحث

ان الإحصاء من النوع Robust هو مقارنة بديلة للإحصاء التقليدي الذي يعطي نتائج ضعيفة عند دراسة معطيات فيها قيم متطرفة (Outliers)

تعتمد الطريقة التقليدية للتقدير في السلاسل الزمنية المالية والتي تعرف ب QMLE على الفرضيات بشكل كبير والتي قد لا تتحقق على أرض الواقع. حيث أنه يفترض أن البيانات موزعة بشكل طبيعي ، إلا أنه عند وجود قيم متطرفة في البيانات يكون لهذه الطريقة كفاءة متدنية جداً. يقدم الإحصاء المتين Robust طرق تحاكي الطرق التقليدية إلا أنها لا تتأثر بالقيم الشاذة أو الخروج الطفيف عن افتراض النموذج للبيانات.

تم في هذه الدراسة محاكاة لعدد من توزيعات الخطأ التي تتراوح بين الذيل الثقيل إلى توزيعات خفيفة الذيل ومن توزيعات متماثلة إلى توزيعات ملتوية لتقييم متانة التوزيعات الثقيلة الذيل في وجود نوعين من القيم المتطرفة Additive و Innovative. وهو ما كشف عن الحاجة إلى استخدام طرق تقدير قوية في تقدير معاملات GARCH في وجود تلك القيم المتطرفة.

ABSTRACT

In financial time series, the conventional fitting procedure (QMLE) suffers from the outlier problem. Estimation of the parameters in GARCH model, can be adversely affected by a single outlier.

simulation studies will not only demonstrate the robustness of this estimate, but will provide evidence as to the utility, efficiency, and validity of this estimate as a robust procedures.

A large Monte Carlo study over error distributions ranging from heavy-tailed to light-tailed distributions and from symmetric distributions to skewed distributions is conducted to evaluate the robustness of heavy tailed distributions in the presence of additive or innovative outliers which revealed the need of robust estimator other than QMLE in estimating GARCH coefficients in the presence of those outliers.

INTRODUCTION

A popular method for estimating the unknown parameters in univariate and multivariate GARCH models is to use the Gaussian likelihood of the innovations and the resulting estimator is called the quasi-maximum likelihood estimator (QMLE).

Traditional methods of estimation and testing, such as the QMLE method, are efficient if the normality assumption of the error distribution and other assumptions about a model are not violated. In particular, stock market data comes along with some skewness, tail weights, outliers and unknown distributions that violates some underlying assumptions for which the estimates from QMLE is efficient.

Most studies employing ARCH and GARCH models document the existence of severe excess kurtosis in the estimated residuals. This non-normality may be due to model misspecifications or outliers. Outliers can affect all the stages of time series analysis: model identification, estimation and forecasting. Outliers may have significant impact on the results of standard methodology for time series analysis, therefore it is important to detect them, estimate their effects and undertake the appropriate corrective actions. A couple of common outliers that may occur in time series data are additive outliers (AO) and innovation outliers (IO).

In this paper we try to prove that the QMLE method is non-robust to the presence of outliers even with fat-tailed and skewed distributions and to analyze the effects of outliers on estimated GARCH volatilities and To study the impact of outliers on the Gaussian QML estimator of GARCH models and some distributions and to study the impact of outliers on the Gaussian QML estimator of GARCH models and some distributions.

(Grané, et al., 2014) studied the impact of additive outliers (level and volatility) on the estimation of MCRRs¹ for short and long trading investment positions. Through simulation and empirical studies, they compared six proposals to reduce the effects of outliers in the estimation of risk measures when using a symmetric volatility model, GARCH(1,1) and

¹MCRR is minimum capital risk requirements

an asymmetric one, GJR(1,1) by (Glosten, et al., 1993) they investigated several Methods to deal with outliers:

[1] Correcting for significant outliers. This approach included the proposal by (Grané, et al., 2010) with hard and soft thresholding, which proceeds by detecting and correcting outliers before estimating these risk measures with symmetric and asymmetric GARCH-type models and also they analyzed the proposal by (Franses, et al., 1999) only for the GARCH model.

[2] Accommodating outliers using fat-tail distributions. This approach consisted in fitting, respectively, t-distributed or skewed t-distributed GARCH-type models directly to the data.

[3] Accounting for outlier effects by the robust estimation of (Muler, et al., 2008).

In the simulation they studied two situations. In the first one, they considered that there was no uncertainty about the data generating process (DGP), whereas in the second one, they analyzed the impact of uncertainty on the MCRRs when the data generating process was unknown. In this case, they considered two sources of uncertainty; functional form and error distribution. They started by simulating return series of different sample sizes. The outliers were placed randomly in the series then they described the considered situations: One isolated ALO of three different magnitudes in simulated series, Two isolated ALOs in simulated series and Patches of three ALOs in series simulated from the considered model. The beginning of the patch was placed randomly in the series. And they also considered one isolated AVO in series simulated from the considered model. They also considered original simulated series (with no outliers) with two cases. In the first The MCRRs were calculated with the same model used to simulate the series. In the second the MCRRs were calculated with a different model from that used to simulate the series. This situation used to analyze the impact of model uncertainty.

Capital risk requirements are estimated for a 1 day investment horizon for the simulated data. They also examined four daily stock market indexes: the FTSE-100 index, the Nasdaq index, the NYSE composite index and the S&P 500 index. For a full evaluation of the results, they performed out-of-

sample conditional tests on the MCRRs calculated. The most important findings in this paper are: Firstly, outliers affected seriously the estimates of MCRRs and the effects depended on outlier magnitudes. Secondly, the detection proposal by (Grané, et al., 2010) with hard thresholding correction almost eliminates the biases on the MCRR estimates, whereas the highest biases are obtained when applying (Muler, et al., 2008) proposal or, in the case of asymmetric models, when considering more complex error distributions. Thirdly, the impact on MCRRs is extremely high in the presence of additive volatility outliers. Fourthly, the results from the study of model uncertainty supported that fitting asymmetric GARCH-type models to simulated series from a GARCH(1,1) led to similar MCRRs (the contrary statement does not hold). Finally, the empirical application and the out-of-sample results for four international stock market indexes indicated that all the methods under study provided statistically equivalent failure rates in estimating the MCRRs for 95% coverage during the last global financial crisis. However, using more general models and more complex distributions generated failure rates closer to 5%.

Generalized ARCH (GARCH) Model

Because ARCH(p) models are difficult to estimate, and because they decay very slowly, Four years after Engel's introduction of the ARCH process, (Bollerslev, 1986), proposed the Generalized ARCH (GARCH) model as a natural solution to the problem with the high ARCH orders. This model is based on an infinite ARCH specification and it allows to dramatically reducing the number of estimated parameters from an infinite number to just a few.

In Bollerslev's GARCH model the conditional variance is a linear function of past squared innovations and earlier calculated conditional variances. The GARCH model has become more popular, because with just a few parameters it can fit data better than a more parameterized ARCH model.

Over the years, the GARCH family has become more efficient in fitting the volatility data as they consist of the second order moment that measures the time-variant of the volatility data. The initial studies by (Engle, 1982) and

(Bollerslev, 1986) turn out to be the better models for volatility (financial) data as the residuals of the data form fatter tailed.

Definition [GARCH(p, q)]

The process $(\varepsilon_t), t \in \mathbb{Z}$, is GARCH (p, q), if:

$$E[\varepsilon_t | \mathcal{F}_{t-1}] = 0, \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$$

The basic model that GARCH type models fit is the following:

$$r_t = \mu_t + \varepsilon_t \quad (1)$$

$$\varepsilon_t = \sigma_t z_t \quad (2)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (3)$$

where $\mu_t = E(r_t | \mathcal{F}_{t-1})$, \mathcal{F}_{t-1} denotes the information set available at time t-1 and z_t is a sequence of i.i.d. random variables with mean 0 and variance 1.

z_t is generally assumed to follow Standard Normal Distribution, Standardized Student-t Distribution or Generalized Error Distribution.

The covariance-stationary condition for GARCH process is:

$$\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 0 \quad (4)$$

The level of persistence is given by $\alpha + \beta$.

The main idea of the GARCH model is that the conditional variance of returns given information available up to time t-1, has an autoregressive structure and is positively correlated to its own recent past and to recent values of the squared returns. This captures the idea of volatility (conditional variance) being persistent which means that large (small) values of r_t^2 are likely to be followed by large (small) values.

Following (Drost, et al., 1993) we can give three interpretations to GARCH(p,q) model:

The strong form GARCH(p,q) process arises when:

$$\text{var}[\varepsilon_t | \mathcal{F}_{t-1}] = \sigma_t^2 \text{ And } Z_t = \varepsilon_t / \sigma_t \text{ is i. i. d.} \quad (5)$$

The semistrong form arises when:

$$\text{var}[\varepsilon_t | \mathcal{F}_{t-1}] = \sigma_t^2 \quad (6)$$

Finally, there is a weak form in which σ_t^2 is defined as a projection on a certain subspace:

$$\mathcal{P}(\varepsilon_t^2 | 1, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots) = \sigma_t^2 \quad (7)$$

Definition [GARCH (1, 1)]

In a GARCH (1,1), the variance (σ_t^2) is a function of an intercept (ω), a shock from the prior period (α) and the variance from the last period (β):

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (8)$$

Outliers in GARCH Models

Additive Outlier; an additive outlier appears as a surprisingly large or small value occurring for a single observation. Subsequent observations are unaffected by an additive outlier. (For example, a data coding error might be identified as an additive outlier). Consecutive additive outliers are typically referred to as additive outlier patches. They are a group of two or more consecutive additive outliers, selecting this outlier type results in the detection of individual additive outliers in addition to patches of them. Additive outliers are the result of adding a value of some magnitude to some of the data points.

Innovational Outlier; an innovational outlier is characterized by an initial impact with effects remaining for a long time over subsequent observations. The influence of the outliers may increase as time proceeds. For stationary series, an innovational outlier affects several observations. For nonstationary series, it may affect every observation starting at a particular series point.

Outliers can cause both the excess kurtosis in the GARCH residuals and the bias in the estimated coefficients and their statistical inferences. GARCH volatilities depend on the unconditional variance, which is a non-

linear function of the parameters. Consequently, they can have larger biases than estimated parameters. One approach to estimate GARCH models in the presence of outliers is to trim iteratively the outliers and fit the model to the remaining data until no more outliers are detected (Franses, et al., 1999). Another approach is to use an estimator that is robust to outliers (Sakata, et al., 1998) , (Park, 2002) and (Peng, et al., 2003)) estimate the GARCH model by minimizing a robust measure of scale of the residuals. (Muler, et al., 2002) and (Muler, et al., 2008) proposed a robust M-estimator that assigns a much lower weight to outliers than the Gaussian ML estimator does.

Distributions in GARCH Model

Normal Distribution

The following log-likelihood function needs to be maximized:

$$\ln(L(\Phi)) = -\frac{1}{2} \sum_{t=1}^n \ln(\sigma_t^2) + \left(\frac{\varepsilon_t}{\sigma_t} \right)^2 \quad (9)$$

Where Φ the vector of the parameters to be estimated

$$(L(\Phi)) = -\sum_{t=1}^n \left(\ln(\sigma_t) + \left(\frac{\nu+1}{2} \right) \ln \left(1 + (\nu-2)^{-1} \left(\frac{\varepsilon_t}{\sigma_t} \right)^2 \right) \right) + T\psi_\nu \quad (10)$$

Where ν is the degrees of freedom and

$$\psi_\nu = \ln \left[((\nu-2)\pi)^{-1/2} \Gamma\left(\frac{\nu}{2}\right)^{-1} \Gamma\left(\frac{\nu+1}{2}\right) \right] \quad (11)$$

Where $\Gamma(\cdot)$ is the gamma distribution. The Student-t distribution incorporates the standard normal distribution as a special case when $\nu = \infty$ and the Cauchy distribution when $\nu = 1$. So, a lower parameter , yields a distribution with fatter tails.

The log-likelihood function for the standard GED is defined as

$$\ln(L(\Phi)) = -\frac{1}{2} \sum_{t=1}^n \left(\ln(\sigma_t^2) + \left| \frac{\varepsilon_t}{\lambda \sigma_t} \right|^v \right) - \phi_v \quad (12)$$

Where λ is defined as :

$$\lambda = [2^{-2/v} \Gamma(1/v) / \Gamma(3/v)]^{1/2} \quad (13)$$

And

$$\phi_v = (1 + v^{-1}) \ln(2) + \ln(\Gamma(1/v)) + \ln\left(\frac{\lambda}{v}\right) \quad (14)$$

GED incorporates the normal ($v = 2$), the Laplace ($v = 1$) and the uniform ($v = \infty$) distribution as special cases, where fatter tails than the normal distribution is obtained for $v < 2$.

The phenomenon of skewness has been recognized in the empirical financial literature for several years and it is well-know the stock return distributions exhibit negative skewness and kurtosis.

(Fernandez, et al., 1998) proposed Skewed Distribution, which allows skewness in any continuous and symmetric distribution by changing the scale at each side of the model

$$f(z|\xi) = \frac{2}{\xi + \frac{1}{\xi}} \left[f(\xi z) H(-z) + f\left(\frac{z}{\xi}\right) H(z) \right] \quad (15)$$

In above model ξ is a shape parameter which is positive and describes the degree of asymmetry. $\xi = 1$ is the symmetric distribution with $f(z|\xi = 1) = f(z)$. $H(z) = (1 + \text{sign}(z)) / 2$ is the Heaviside unit step function.

If a skewed distribution functions with zero mean and unit variance, it is called Standardized Skewed Distribution.

The probability function $f(z|\xi)$ of a Standardized Skewed Distribution can be written as:

$$f(z|\xi\theta) = \frac{2\sigma}{\xi + \frac{1}{\xi}} f\left(z_{u_\xi \sigma_\xi} \middle| \theta\right) \quad (16)$$

Where $z_{\mu_\xi \sigma_\xi} = \xi^{sign(\sigma_\xi z + u_\xi)} (\sigma_\xi z + u_\xi)$, $f(\cdot)$ is any standardized symmetric distribution function, like the Standard Student-t Distribution, the Standard Generalized Error Distribution or Standard Normal Distribution.

In above equation, μ is location parameter, σ is the standard deviation parameter, ξ is a shape parameter which model the skewness, θ is an optional set of shape parameter that models higher moments of even order like ν in GED and Student-t Distribution.

Skewed-Normal Distribution

The probability density function of the Skewed-Normal Distribution with parameter α is given by

$$f(x) = 2\phi(x)\Phi(\xi x) \quad (17)$$

Where ξ is a fixed parameter called the shape parameter that control the different shape of density function.

$$\begin{aligned} \phi(x) &= \exp(-x^2/2)/\sqrt{2\pi} \\ \Phi(\xi x) &= \int_{-\infty}^{\xi x} \phi(t) dt \end{aligned}$$

Skewed-Student-t Distribution

Based on the Standard Skewed Distribution, the Skewed-Student-t Distribution pdf is:

$$f(z_t|\nu, \xi) = \begin{cases} \frac{2\sigma}{\xi + \xi^{-1}} g[\xi(sz_t + \mu)|\nu] \cdots \cdots \cdots z_t < -\frac{\mu}{\sigma} \\ \frac{2\sigma}{\xi + \xi^{-1}} g[(sz_t + \mu)\xi|\nu] \cdots \cdots \cdots z_t \geq -\frac{\mu}{\sigma} \end{cases} \quad (18)$$

$$z_t \sim SKST(0,1, \xi, \nu)$$

ξ is non-symmetry parameter, ν test the fat tail, μ and σ^2 are the mean and variance of non-Standard Student-t Distribution.

Standardized Skewed-Generalized Error Distribution

The density function of the Standardized Skewed-Generalized Error Distribution is

$$f(z_t|\nu, \xi) = \nu(2\theta * \Gamma(1/\nu))^{-1} * \exp\left(-\frac{|z_t-\delta|^\nu}{[1-\text{sign}(z_t-\delta)\xi]^\nu\theta^\nu}\right) \quad (19)$$

Where

$$\theta = \Gamma(1/\nu)^{0.5}\Gamma(3/\nu)^{-0.5}S(\xi)^{-1}$$

$$\delta = 2\xi * AS(\xi)^{-1}$$

$$S(\xi) = \sqrt{1 + 3\xi^2 - 4A^2\xi^2}$$

$$A = \Gamma(2/\nu)\Gamma(1/\nu)^{-0.5}\Gamma(3/\nu)^{-0.5}$$

Where the shape parameter ν controls the height and fat-tail of the density function with constraint $\nu > 0$, while ξ is a skewness parameter of the density with $-1 < \xi < 1$. In the case of positive skewness, the density function skews toward to the right, vice versa. When $\nu = 2$, $\lambda = 0$, the SGED distribution turns out to be the standard normal distribution.

Quasi-Maximum Likelihood Estimation (QMLE)

As it is stated in (Huang, et al., 2008) the quasi-maximum likelihood (QML) method is particularly relevant for GARCH models because it provides consistent and asymptotically normal estimators for strictly stationary GARCH processes under mild regularity conditions, but with no moment assumptions on the observed process.

Conditional Quasi-Likelihood is given by

$$\ln(L(\Phi)) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \quad (20)$$

The GARCH (1,1) shown in Equation (2.14) can be expressed as

$$\begin{aligned} \varepsilon_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \varepsilon_t^2 - \sigma_t^2 \\ \varepsilon_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 - \beta\varepsilon_{t-1}^2 + \beta\varepsilon_{t-1}^2 + \varepsilon_t^2 - \sigma_t^2 \\ \varepsilon_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\varepsilon_{t-1}^2 + \beta(\sigma_{t-1}^2 - \varepsilon_{t-1}^2) + \varepsilon_t^2 - \sigma_t^2 \\ \varepsilon_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\varepsilon_{t-1}^2 - \beta v_{t-1} + v_t \\ \varepsilon_t^2 &= \omega + (\alpha + \beta) \varepsilon_{t-1}^2 - \beta v_{t-1} + v_t \end{aligned}$$

Where $v_t = \varepsilon_t^2 - \sigma_t^2$ and the conditional mean of v_t is zero $E(v_t|\mathcal{F}_{t-1}) = 0$. For deriving the unconditional variances of ε_t , the conditional variance for the GARCH (1, 1) return process is given by

$$\begin{aligned} \text{var}(\varepsilon_t) &= E(\varepsilon_t^2) - (E(\varepsilon_t))^2 = E(\varepsilon_t^2) \\ \text{var}(\varepsilon_t) &= E[\omega + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta v_{t-1} + v_t] \\ \text{var}(\varepsilon_t) &= \omega + E[(\alpha + \beta)\varepsilon_{t-1}^2] \\ \text{var}(\varepsilon_t) &= \frac{\omega}{1 - (\alpha + \beta)} \end{aligned}$$

since $E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2)$ The requirements for stationarity are such that $1 - \alpha - \beta > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\omega > 0$.

The estimation of parameters of the GARCH (1, 1) model is carried in a similar fashion as (Bollerslev, 1986) suggested that the unconditional variance for ε_t should be considered as a starting value for variance. That is,

$$E(\varepsilon_t^2) = \frac{\omega}{1 - \alpha - \beta} \quad (21)$$

Under the normality assumption, the likelihood function of the GARCH (1, 1) model is expressed as

$$\begin{aligned} f(\varepsilon_1, \dots, \varepsilon_T, \sigma_1^2, \dots, \sigma_T^2) \\ = f(\varepsilon_T, \sigma_T^2 | \mathcal{F}_{T-1}) f(\varepsilon_{T-1}, \sigma_{T-1}^2 | \mathcal{F}_{T-2}) \cdots f(\varepsilon_2, \sigma_2^2 | \mathcal{F}_1) f(\varepsilon_1, \sigma_1^2 | \theta) \end{aligned}$$

where $\theta = [\omega, \alpha, \beta]'$ is the vector of unknown parameters and $f(\varepsilon_1, \sigma_1^2 | \theta)$ is a probability density function. Since $f(\varepsilon_1, \sigma_1^2 | \theta)$ is complicated, it is commonly dropped from the conditional likelihood function, especially

when the sample size is sufficiently large. Hence, the conditional likelihood function is expressed as follows:

$$f(\varepsilon_2, \dots, \varepsilon_T, \sigma_2^2, \dots, \sigma_T^2 | \theta, \varepsilon_1^2, \sigma_1^2) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\} \quad (22)$$

And is considered to estimate θ . The maximum likelihood estimates are obtained by directly maximizing this expression or, equivalently the log-likelihood function given by:

$$L(\theta | \varepsilon_1, \sigma_1^2) = \sum_{t=2}^T L_t(\theta) \quad (23)$$

where $L_t(\theta)$ is the log-likelihood function of observation t and is calculated as

$$L_t(\theta) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2} \quad (24)$$

Studying the Effect of Additive and Innovative Outliers on GARCH (1,1) Model

Financial time series generally have the characteristics of being leptokurtotic, fat-tailed with skewness and volatility clustering. So it violates the Normal distribution assumption for the errors. Most of Literatures suppose that the errors follow Normal distribution, but for many financial series, it cannot describe the character of the fat tails, so using other distribution to describe fat-tail distribution is needed. In this simulation a comparison of GARCH(1,1) Model based on Different Distributions is done based on Normal Distribution, Student-t Distribution, Generalized Error Distribution and their skewed version.

First a simulation of uncontaminated series is generated with specifications that accounts for heavy tails and skewness that are characters of financial time series and then are contaminated with outliers with different numbers and magnitudes to study the effect of outliers either additive or innovative on those series.

The initial parameters of the GARCH(1,1) models are set within the stationary region and selected to represent three cases of volatility persistence. The persistence indicates how fast the volatility decays and regain normalcy. A weak persistence shows quick decay of volatility. A strong persistence shows slow decay of volatility. α and β are coefficients of conditional squared residuals and conditional variance. $(\alpha + \beta)$ determines the volatility persistence. The sample sizes N are varied as 250, 500 and 1000. Large sample sizes (1000) are chosen to approximately represent heavy set of real life financial data and also to observe the structure more accurately.

The main purpose of the this simulation is to investigate effects of additive outliers and innovative outliers on the performance of the GARCH (1,1) model with various error distribution (Normal Distribution, Student-t Distribution, Generalized Error Distribution) and their skewed version for a total of 36 different specifications for uncontaminated series and series that is contaminated with outliers of different types (AO and IO) with different numbers (1,2%,5%) and magnitudes ($5\sigma, 10\sigma, 15\sigma$) and to choose the best distribution for the innovations for such series.

The second purpose of this simulation is to study the robustness of various distributional assumption in the previously stated cases. From Standardized Residuals Diagnosis it can be concluded that the estimates of the underlying volatility in a GARCH(1,1) model can be seriously affected by outliers through their effect on the estimates of the parameters and that fat tailed distributions didn't provide robust estimate of GARCH (1,1) parameters.

A 1000 replications is used in each model to study the performance. All the simulations are performed using R.

Data Generating Process (DGP)

The Monte Carlo experiment is set up using the GARCH(1,1) DGP:

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

Parameter values

1. High persistence $\omega = 0.01$ $\alpha = 0.09$ $\beta = 0.90$

2. Medium persistence $\omega = 0.10$ $\alpha = 0.10$ $\beta = 0.80$
3. Low persistence $\omega = 0.10$ $\alpha = 0.20$ $\beta = 0.60$

The distributions

The error distribution $\varepsilon_t = \sigma_t z_t$ where z_t is assumed to follow different distributions: (ν indicates degrees of freedom while ξ is the skewness parameter)

1. Normal distribution (N).
2. Student t distribution (std) $\nu = 3$.
3. Student t distribution (std) $\nu = 5$.
4. Student t distribution (std) $\nu = 7$.
5. Student t distribution (std) $\nu = 9$.
6. Generalized Error Distribution (GED) $\nu = 1$.
7. Generalized Error Distribution (GED) $\nu = 1.5$.
8. Generalized Error Distribution (GED) $\nu = 2$.
9. Generalized Error Distribution (GED) $\nu = 2.3$.
10. Skewed normal distribution (SN) $\xi = 0.5$.
11. Skewed normal distribution (SN) $\xi = 0.8$.
12. Skewed normal distribution (SN) $\xi = 0.9$.
13. Skewed Student t distribution (sstd) $\nu = 3, \xi = 0.5$.
14. Skewed Student t distribution (sstd) $\nu = 3, \xi = 0.8$.
15. Skewed Student t distribution (sstd) $\nu = 3, \xi = 0.9$.
16. Skewed Student t distribution (sstd) $\nu = 5, \xi = 0.5$.
17. Skewed Student t distribution (sstd) $\nu = 5, \xi = 0.8$.
18. Skewed Student t distribution (sstd) $\nu = 5, \xi = 0.9$.
19. Skewed Student t distribution (sstd) $\nu = 7, \xi = 0.5$.
20. Skewed Student t distribution (sstd) $\nu = 7, \xi = 0.8$.
21. Skewed Student t distribution (sstd) $\nu = 7, \xi = 0.9$.
22. Skewed Student t distribution (sstd) $\nu = 9, \xi = 0.5$.
23. Skewed Student t distribution (sstd) $\nu = 9, \xi = 0.8$.
24. Skewed Student t distribution (sstd) $\nu = 9, \xi = 0.9$.
25. Skewed Generalized Error Distribution (SGED) $\nu = 1, \xi = 0.5$.
26. Skewed Generalized Error Distribution (SGED) $\nu = 1, \xi = 0.8$.
27. Skewed Generalized Error Distribution (SGED) $\nu = 1, \xi = 0.9$.

28. Skewed Generalized Error Distribution (SGED) $\nu = 1.5, \xi = 0.5$.
29. Skewed Generalized Error Distribution (SGED) $\nu = 1.5, \xi = 0.8$.
30. Skewed Generalized Error Distribution (SGED) $\nu = 1.5, \xi = 0.9$.
31. Skewed Generalized Error Distribution (SGED) $\nu = 2, \xi = 0.5$.
32. Skewed Generalized Error Distribution (SGED) $\nu = 2, \xi = 0.8$.
33. Skewed Generalized Error Distribution (SGED) $\nu = 2, \xi = 0.9$.
34. Skewed Generalized Error Distribution (SGED) $\nu = 2.3, \xi = 0.5$.
35. Skewed Generalized Error Distribution (SGED) $\nu = 2.3, \xi = 0.8$.
36. Skewed Generalized Error Distribution (SGED) $\nu = 2.3, \xi = 0.9$.

Model Selection Criteria

Various evaluation measures are employed to evaluate the parameter estimates contaminated and uncontaminated series.

According to (Akaike, 1973), a model should be evaluated on the basis of good results when it is used for prediction. The author suggested a method for evaluating model in terms of Kullback-Leibler information, which is based on the concept of closeness between generic distribution $g(x)$ defined by the model and the true distribution $f(x)$, besides the more commonly used method of simply minimizing the prediction error. The Kullback-Leibler information $I(f, g)$ is the information lost when model $g(x)$ is used to approximate $f(x)$; this is defined for continuous functions as the integral

$$I(f, g) = \int f(x) \ln \left[\frac{f(x)}{g(x|\theta)} \right] dx$$

There will be two types of tests performed on the series to see which distributions is considered most fitting on the contaminated GARCH time-series. The tests are :

Akaike Information Criteria (AIC):

The Akaike Information Criteria (AIC) has commonly been used and significantly known method in the model selection for decades. The AIC compares models from the perspective of information entropy, as measured by Kullback-Leibler divergence. The AIC for a given model is:

$$AIC = -2 \ln L + 2k$$

L is the maximized value of the likelihood function and k is the number of free parameters used in the model. When comparing AIC values for multiple models, smaller values of the criteria are better.

Bayesian Information Criteria (BIC):

(Schwarz, 1978) developed a consistent criteria for models defined in terms of their posterior probability (Bayesian approach). The main difference between the AIC and BIC is that BIC penalizes models more when number of parameters increases. The BIC, also known as Schwarz information criteria, compares models from the perspective of decision theory, as measured by expected loss. The BIC for a given model is :

$$\text{BIC} = -2 \ln L + k \ln T$$

where T is the number of observations. The preferred model will show the lowest values of this criteria.

The verification of the robustness in the estimation with various distributional assumption is achieved through studying the Quantile-Quantile Plot of Standardized Residuals.

QQ plot is a way to visually examine if empirical data fits the reference or hypothesized theoretical distribution. The process graphs the quantiles at regular confidence intervals for the empirical distribution against the theoretical distribution. As an example, if both the empirical and theoretical data are drawn from the same distribution, then the median (confidence level = 50%) of the empirical distribution would plot very close to zero, while the median of the theoretical distribution would plot exactly at zero. Continuing in this fashion for other quantiles (40%, 60%, and so on) will map out a function. If the two distributions are very similar, the resulting Q-Q plot will be linear.

The QQ-plots for standardized residuals show the estimated innovations against the sample quantiles from a normal distribution. If the plotted scatters fall on the straight line, the estimated innovations follow a normal distribution.

Conclusions and Recommendations

Overall results from this study provide evidence to show that skewed fat tail distributions are more appropriate for the financial time series.

1. It is important to incorporate fat-tailed innovations in fitting all the asset prices series to account for leptokurtosis into GARCH processes
2. The performance of the considered types of fat-tailed distributions and their skewed versions in fitting the financial data is quite different according to degree of persistence and sample size and type of outlier included, and the skewed student -t distribution dominates the other type of distribution.

Recommendations

GARCH models with the skewed student -t distribution perform very well and therefore recommended for financial time series in QMLE estimation.

References

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* , 31 (3), 307-327.
- Carnero, P. a. (2007). EFFECTS OF OUTLIERS ON THE IDENTIFICATION AND ESTIMATION OF GARCH MODELS. *Time Series Analysis* , 28 (4), 471–497.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* , 50 (4), 987-1007.
- Fernandez, C., & Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* , 93 (441), 359-371.
- Fox, A. (1972). Outliers in time series. *J. R. Stat. , B* (34), 350–363.
- Franses, P., & Ghijssels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting* , 15, 1-9.
- Grané, A., & Veiga, H. (2014). Outliers, GARCH-type models and risk measures: A comparison of several approaches. *Journal of Empirical Finance* , 26, 26-40.
- Huang, D., Wang, H., & Yao, q. (2008). Estimating GARCH Models: When to Use What? *Econometric journal* , 11, 27-38.
- Muler, N., & Yohai, V. (2002). Robust estimates for ARCH processes. *Journal of Time Series Analysis* , 23 (3), 341-375.
- Park, B. (2002). An outlier robust GARCH model and forecasting volatility of exchange rate returns. *Journal of forecasting* , 21, 381-393.
- Sakata, S., & White, H. (1998). High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica* , 66 (3), 529-568.