Economics Letters 162 (2018) 93–97

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Robust maximum entropy test for GARCH models based on a minimum density power divergence estimator

Byungsoo Kim

Department of Statistics, Yeungnam University, Kyungsan, 38541, Republic of Korea

HIGHLIGHTS

- We develop a robust maximum entropy test for the normality of GARCH models.
- As a robust estimator, we employ the minimum density power divergence estimator.
- We derive the limiting null distribution of the test statistics.
- A bootstrap method is also discussed.
- Our test overcomes size distortions remarkably well in the presence of outliers.

ARTICLE INFO

Article history:
Received 9 May 2017
Received in revised form 2 November 2017
Accepted 3 November 2017
Available online 10 November 2017

JEL classification:
C12
C13
C22

Keywords:
Entropy-based goodness-of-fit test
Normality test
GARCH models
Minimum density power divergence estimator
Parametric bootstrap method

ABSTRACT

The maximum entropy test, as designed for examining goodness-of-fit with a non-robust estimator such as the maximum likelihood estimator, can suffer from severe size distortions when the data are contaminated by outliers. The objective of this study is to develop a robust maximum entropy test for the normality of GARCH models. We construct the test statistic based on the minimum density power divergence estimator and verify its limiting null distribution. A bootstrap method is also discussed, and its performance is evaluated through simulations. According to the simulation results, the proposed test can successfully achieve reasonable sizes in the presence of outliers.

1. Introduction

For a probability density function f, the Boltzmann–Shannon entropy is defined by

$$H(f) = -\int_{-\infty}^{\infty} f(x)\log f(x)dx. \quad (1)$$

Forte and Hughes (1988) proposed a discrete analogue of (1) of the form $-\sum p_i \log(p_i/(x_i - x_{i-1}))$, where $p_i = P(x_{i-1} < X < x_i) = \int_{x_{i-1}}^{x_i} f(x)dx,\ i = 1, \dots, n-1$ and $a = x_0 < \cdots < x_n = b$. Based on this quantity, Lee et al. (2011) proposed a maximum entropy test in independent and identically distributed (iid) random variables and demonstrated that the test outperforms other existing goodness-of-fit (gof) tests.

The gof tests play an important role in matching given datasets with the best fitted ones among the candidate distribution families. In time series analysis, testing for normality is a crucial issue. In particular, the normality test for GARCH innovations is important because the Gaussian quasi-maximum likelihood estimator (QMLE) of GARCH parameters, which is a widely used estimation technique for GARCH models, becomes less efficient if the innovations are non-Gaussian. Further, normality tests are vital in many applications, such as calculating the conditional value-at-risk. As the normality of innovations does not necessarily imply that of the GARCH model itself, the normality test should be applied to the innovations. Recently, Lee et al. (2015) developed a maximum entropy gof test for GARCH(1,1) innovations based on QMLE. However, as is widely recognized in the literature, the maximum likelihood method is unduly influenced by outliers. Hence, the maximum entropy normality test for GARCH innovations based on

E-mail address: bkim@yu.ac.kr.

https://doi.org/10.1016/j.econlet.2017.11.003
0165-1765/

QMLE can suffer from severe size distortions when the data are contaminated by outliers. To cope with such a defect, we propose a maximum entropy test based on a robust estimator instead of the QMLE, and compare the performance of the proposed test with that of the QMLE-based entropy test. For this task, we employ the minimum density power divergence estimator (MDPDE) proposed by Basu et al. (1998, BHHJ hereafter) as a robust estimator. BHHJ developed an estimation procedure that is robust against outliers by minimizing the divergence between two density functions, which is called the density power divergence. Compared with previous divergence-based estimation methods, BHHJ's approach has the advantage that it does not require any smoothing methods to estimate the true density function of the data. Owing to this advantage, the MDPDE has been applied to various models: for example, Lee and Song (2009) and Kim and Lee (2013) recently studied the MDPDE for GARCH models and the covariance matrix of multivariate times series. BHHJ showed that the MDPDE possesses strong robust properties with little loss in asymptotic efficiency relative to the MLE. Thus, the MDPDE can be regarded as a good alternative to the MLE in terms of both robustness and asymptotic efficiency.

The remainder of this paper is organized as follows. In Section 2, we construct the MDPDE for GARCH(1,1) parameters. In Section 3, we introduce the test statistic of the maximum entropy test and its limiting null distribution in iid samples, and apply this test to GARCH(1,1) models with the MDPDE. In Section 4, we conduct a simulation study to evaluate the capabilities of the proposed test.

2. MDPDE for GARCH models

For two given density functions f and g, their density power divergence is defined by

$$d_\lambda(g,f) := \begin{cases} \int \{f^{1+\lambda}(y) - (1+\frac{1}{\lambda})g(y)f^\lambda(y) + \frac{1}{\lambda}g^{1+\lambda}(y)\}dy, & \lambda > 0, \\ \int g(y)\{\log g(y) - \log f(y)\}dy, & \lambda = 0. \end{cases}$$

Then, for a parametric family $\{F_\theta\}$, indexed by some unknown parameter $\theta \in \Theta$ and possessing a density of $\{f_\theta\}$, and a distribution G with density g, the minimum density power divergence functional $T_\lambda(G)$ is defined by

$$d_\lambda(g, f_{T_\lambda(G)}) = \min_{\theta\in\Theta} d_\lambda(g, f_\theta).$$

In particular, if $G = F_{\theta_0} \in \{F_\theta\}$, $T_\lambda(G) = \theta_0$. Based on this, given a random sample $X_1, \ldots, X_n$ with unknown density g, the MDPDE is defined as

$$\hat\theta_{\lambda,n} = \underset{\theta\in\Theta}{\operatorname{argmin}} H_{\lambda,n}(\theta),$$

where $H_{\lambda,n}(\theta) = \frac{1}{n}\sum_{t=1}^n V_\lambda(\theta; X_t)$ and

$$V_\lambda(\theta; X_t) = \begin{cases} \int f_\theta^{1+\lambda}(y)dy - \left(1+\frac{1}{\lambda}\right) f_\theta^\lambda(X_t), & \lambda > 0, \\ -\log f_\theta(X_t), & \lambda = 0. \end{cases}$$

When $\lambda = 0$ and 1, the MDPDE is the same as the MLE and L_2-distance estimator, respectively. BHHJ showed that $\hat\theta_{\lambda,n}$ is consistent for $T_\lambda(G)$ and asymptotically normal. They also demonstrated that the estimator is robust against outliers, but still highly efficient when the true distribution belongs to the parametric family $\{F_\theta\}$ and λ is close to 0. To apply the above procedure to GARCH models, we must define the conditional version of the MDPDE. As the true conditional distribution of the time series X_t given $\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is a σ-field generated by $\{X_{t-1}, X_{t-2}, \ldots\}$, is usually unknown in practice, we consider a candidate parametric conditional distribution family $\{f_\theta(\cdot|\mathcal{F}_{t-1})\}$ indexed by the parameter θ to play the role of the true conditional distribution. Then, the MDPDE is defined by

$$\hat\theta_{\lambda,n} = \underset{\theta\in\Theta}{\operatorname{argmin}} H_{\lambda,n}(\theta),$$

where $H_{\lambda,n}(\theta) = \frac{1}{n}\sum_{t=1}^n V_\lambda(\theta; \mathcal{F}_{t-1}, X_t)$ and

$$V_\lambda(\theta; \mathcal{F}_{t-1}, X_t) = \begin{cases} \int f_\theta^{1+\lambda}(y|\mathcal{F}_{t-1})dy - \left(1+\frac{1}{\lambda}\right) f_\theta^\lambda(X_t|\mathcal{F}_{t-1}), & \lambda > 0, \\ -\log f_\theta(X_t|\mathcal{F}_{t-1}), & \lambda = 0. \end{cases} \quad (2)$$

Now, suppose that $X_1, \ldots, X_n$ are observed from the GARCH(1,1) model:

$$X_t = \sigma_t(\theta)\epsilon_t,$$
$$\sigma_t^2(\theta) = \omega + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2(\theta), \quad (3)$$

where ϵ_t are iid random variables with zero mean and unit variance and $\theta = (\omega, \alpha, \beta)' \in \Theta \subset (0,\infty) \times [0,\infty)^2$ with $\alpha + \beta < 1$. Let θ_0 denote the true value of θ. As $\epsilon_t \sim N(0,1)$ implies $X_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2(\theta_0))$ and the purpose of this study is to develop a robust normality test for GARCH innovations that avoids size distortions in the presence of outliers, we employ the normal family $\{N(0,\sigma_t^2(\theta)),\ \theta \in \Theta\}$ as the parametric family $\{f_\theta(\cdot|\mathcal{F}_{t-1})\}$ in (2). Then, from (2), we can define the MDPDE for GARCH parameters $\hat\theta_{\lambda,n} = (\hat\omega_{\lambda,n}, \hat\alpha_{\lambda,n}, \hat\beta_{\lambda,n})'$ as

$$\hat\theta_{\lambda,n} = \underset{\theta\in\Theta}{\operatorname{argmin}} \tilde H_{\lambda,n}(\theta),$$

where

$$\tilde H_{\lambda,n}(\theta) = \begin{cases} \frac{1}{n}\sum_{t=1}^n \left(\frac{1}{\sqrt{\tilde\sigma_t^2}}\right)^\lambda \left\{\frac{1}{\sqrt{1+\lambda}} - \left(1+\frac{1}{\lambda}\right)\exp\left(-\frac{\lambda X_t^2}{2\tilde\sigma_t^2}\right)\right\}, \\ \lambda > 0, \\ \frac{1}{n}\sum_{t=1}^n \left(\frac{X_t^2}{\tilde\sigma_t^2} + \log\tilde\sigma_t^2\right), \quad \lambda = 0, \end{cases}$$

and $\tilde\sigma_t^2$ are defined recursively by $\tilde\sigma_t^2 = \tilde\sigma_t^2(\theta) = \omega + \alpha X_{t-1}^2 + \beta\tilde\sigma_{t-1}^2$ with initial values X_0^2 and $\tilde\sigma_0^2$. Following the suggestion of Francq and Zakoïan (2004), we choose $X_0^2 = \tilde\sigma_0^2 = X_1^2$ in this study.

3. Maximum entropy test

3.1. Maximum entropy test for an iid sample

In this section, we briefly review the maximum entropy test for iid random variables. Let Y_i, $i = 1, 2, \ldots, n$ be a random sample from a distribution with an unknown distribution function G. Suppose that we wish to test the hypotheses:

$$H_0 : G = G_0 \quad vs. \quad H_1 : G \neq G_0. \quad (4)$$

To deal with this problem, Lee et al. (2011) suggested the following generalization of the entropy defined by Forte and Hughes (1988):

$$S^w(G) = -\sum_{i=1}^m w_i(G(s_i) - G(s_{i-1}))\log\left(\frac{G(s_i) - G(s_{i-1})}{s_i - s_{i-1}}\right),$$

where the w are appropriate weights with $0 \le w_i \le 1$ and $\sum_{i=1}^m w_i = 1$, m is the number of disjoint intervals for partitioning the data range, and $-\infty < a \le s_0 \le \cdots \le s_m \le b < \infty$ are preassigned partition points. As $S^w(G_0) = 0$ when G_0 is uniformly distributed in [0, 1], the gof test can be reduced to a uniform test based on the probability integral transform $G_0(Y_i)$, denoted by U_i. The role of weights is important for the test. Lee et al. (2011)