Tests of fit for symmetric variance gamma distributions

Fragiadakis Kostas
UADPhilEcon, National and Kapodistrian University of Athens, 14 Euripidou Street, 105 59 Athens, Greece.

Keywords: Variance Gamma Distribution, Characteristic Function, EM - type algorithm, Goodness of Fit

Abstract

New goodness–of–fit tests for the family of symmetric Variance Gamma distributions are constructed. A Generalized EM type algorithm for the parameters estimation have been used. The proposed tests are based on a weighted integral incorporating the empirical characteristic function.

1. Introduction

The aim of this paper is to provide goodness-of-fit tests for the symmetric normal variance gamma distribution (SNVG). The NVG distribution is defined as a mixture of a normal distribution with a Gamma distribution. Specifically, if $V \sim \Gamma(1, \lambda)$ distribution where $\Gamma(\alpha, \beta)$ denotes the Gamma distribution with density

$$f(x) = \frac{x^{\alpha-1}\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta x}, \qquad \alpha, \beta, x > 0.$$

and if $Z|V \sim N(0, 2V)$, where $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, then the distribution of Y where $Y = \delta + cZ$ is the SNVG distribution with density function

$$f(y) = \frac{1}{c\Gamma(\lambda)\sqrt{\pi}} \left(\frac{y-\delta}{2c}\right)^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\frac{y-\delta}{c}\right)$$

where $K_r(x)$ denotes the modified Bessel function of order r evaluated at x.

Therefore, the SNVG distribution is a three - parameter model, denoted by $SNVG(\delta, c, \lambda)$ which depends on a location parameter $\delta \in \mathbf{R}$, a scale parameter $c > 0$, and a shape parameter $\lambda > 0$.

Suppose that on the basis of independent copies $X_1, X_2, \ldots, X_n$, of a random variable X we wish to test the null hypothesis

H_0: The law of X is $SNVG(\delta, c, \lambda)$ for some $\delta \in \mathbf{R}$, $c > 0$ and $\lambda > 0$.

The motivation for considering the SNVG is that this distribution, due to the heavier tails with respect to normality, becomes a competitive alternative choice of model for applications in Economics, and particularly for modelling financial data.

2. Estimation of Parameters

A critical issue is the estimation of the parameters. Given a random sample $X_1, X_2, \ldots, X_n$, the log likelihood take the form

$$\ell(\delta, c, \lambda) = -n \log(c\Gamma(\lambda)\sqrt{\pi}) + \left(\lambda - \frac{1}{2}\right) \sum_{i=1}^{n} \log\left(\frac{x_i - \delta}{2c}\right) +$$

$$+ \sum_{i=1}^{n} \log\left(K_{\lambda - \frac{1}{2}}\left(\frac{x_i - \delta}{c}\right)\right)$$

which is rather complicated to be maximized. Instead a Generalized EM approach will be used.

Firstly an EM type algorithm is described. Based on the mixture representation we need to augment the observed data $x_1, x_2, \ldots, x_n$, and the unobserved data $v_1, v_2, \ldots, v_n$. At the E – step we need the conditional expectations of some functions of v_i. The conditional distribution of $V_i|X_i = x_i$ can be easily seen to be a Generalized Inverse Gaussian distribution of the form

$$V|X = x \sim GIG\left(\lambda - \frac{1}{2}, \frac{x_i - \delta}{c\sqrt{2}}, \sqrt{2}\right)$$

We know that if $X \sim GIG(\lambda, \delta, \gamma)$ distribution it holds that

$$E(X^r) = \left(\frac{\delta}{\gamma}\right)^r \frac{K_{\lambda+r}(\delta\gamma)}{K_\lambda(\delta\gamma)}$$

The log likelihood of the complete data (X_i, V_i), $i = 1, 2, \ldots, n$ factorizes in two parts and hence one can derive that

$$c^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \delta)^2}{2v_i}, \qquad \delta = \frac{\sum_{i=1}^{n} \frac{x_i}{v_i}}{\sum_{i=1}^{n} \frac{1}{v_i}}$$

and λ is the solution of the equation

$$\Psi(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \log(v_i)$$

This implies that the conditional expectations needed for the M – step have the form $E(V_i^{-1}|X_i)$ and $E(\log V_i|X_i)$. Despite the fact that

$$E(V_i^{-1}|X_i) = \frac{2c}{x_i - \delta} \frac{K_{\lambda - \frac{3}{2}}\left(\frac{x_i - \delta}{c}\right)}{K_{\lambda - \frac{1}{2}}\left(\frac{x_i - \delta}{c}\right)}$$

unfortunately the conditional expectation $E(\log V_i|X_i)$ cannot be written in a useful closed form formula and hence it is hard to be derived.

There are two distinct solutions for this problem. The first one is to estimate this expectation using Monte Carlo, resulting to a MCEM algorithm and the second one is to use a GEM (Generalized EM). The second approach will be used. The idea is instead of solving $\Psi(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\log(v_i)$ we just seek a λ that increases the log–likelihood.

The algorithm has the following steps

E – Step

Calculate with the current estimates

$$s_i = E(V_i^{-1}|X_i) = \frac{2c}{x_i - \delta}\frac{K_{\lambda-\frac{3}{2}}\left(\frac{x_i-\delta}{c}\right)}{K_{\lambda-\frac{1}{2}}\left(\frac{x_i-\delta}{c}\right)}$$

M – Step

Update the parameters using

$$c^2 = \frac{1}{2n}\sum_{i=1}^{n} s_i(x_i - \delta)^2, \qquad \delta = \frac{\sum_{i=1}^{n} x_i s_i}{\sum_{i=1}^{n} s_i}$$

Find a new λ by a grid search in the neighborhood of the current value.

It suffices to find a value for λ that improves the log – likelihood and not necessarily the maximum one. Since for given λ the other two estimates improve the log–likelihood, a new λ that provides better log–likelihood ensures the monotonic property of the EM.

3. Test Statistics

At this section we study a new family of omnibus tests of H_0 based on the empirical characteristic function (CF). Despite the fact that the density function of X is complicated, the CF, $\phi(t) = \mathbf{E}(e^{itX})$ of X is simply

$$\phi(t;\delta,c,\lambda) = e^{i\delta t}(1 + c^2t^2)^{-\lambda}. \tag{1}$$

It is natural to construct a test statistic based on the standardized data i.e. $\delta = 0$ and $c = 1$. Hence, the characteristic function of the standardized data is

$$\phi(t;\lambda) = \frac{1}{(1+t^2)^{\lambda}}$$

By taking the first derivative we have

$$(1 + t^2)\phi'(t) + 2\lambda t\phi(t) = 0 \tag{2}$$

Therefore if $Y_1, Y_2, \ldots, Y_n$ follows the SNVG distribution, equation $D(t;\lambda) = 0$ where $D(t;\lambda)(1+t^2)\phi'(t) + 2\lambda t\phi(t)$ must be satisfied for the standardized data

$X_j = \frac{Y_j - \hat{\delta}_n}{\hat{c}_n}$, $j = 1, 2, \ldots, n$. Taking into account the empirical analogue of CF we have

$$D_n(t; \hat{\lambda}_n) = (1+t^2)\phi_n'(t) + 2\hat{\lambda}_n t \phi_n(t) \qquad \text{where} \qquad \phi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j}$$

Specifically I suggest to reject the null hypothesis H_0 for large values of

$$\hat{T}_{n,w} = n \int_{-\infty}^{\infty} |D_n(t; \hat{\lambda}_n)|^2 w(t) dt, \tag{3}$$

with $w(t)$ denoting a non–negative weight function. Notice that the test statistic remains invariant under the linear transformation $X \mapsto \delta + cX$, for each $\delta \in \mathbf{R}$ and $c > 0$. From (3) we have by straightforward algebra

$$\hat{T}_{n,w} \frac{1}{n} \sum_{j,k=1}^{n} \Big[4\hat{\lambda}_n^2 I_c^{(2)}(X_j - X_k) + X_j X_k \Big(I_c^{(4)}(X_j - X_k) + 2I_c^{(2)}(X_j - X_k) +$$

$$+ I_c^{(0)}(X_j - X_k) \Big) + 2\hat{\lambda}_n (X_j - X_k) \Big(I_s^{(3)}(X_j + X_k) - I_s^{(3)}(X_j - X_k) +$$

$$+ I_s^{(1)}(X_j + X_k) - I_s^{(1)}(X_j - X_k) \Big) \Big]$$

where

$$I_c^{(m)}(b) \int_{-\infty}^{+\infty} t^m \cos(bt) w(t) dt, \quad m = 0, 2, 4,$$

$$I_s^{(p)}(b) = \int_{-\infty}^{+\infty} t^p \sin(bt) w(t) dt, \quad p = 1, 3.$$

Although theoretical properties of the test statistic remain qualitatively invariant, provided that $w(t)$ satisfies some general conditions, particular appeal lies with weight functions that render the test statistic in a closed formula suitable for computer implementation.

4. Bootstrap Procedure

In this section a bootstrap procedure for the new test given by (3) is proposed. To actually implement the test critical points are required. However, the null distribution of the test statistic depends on the value of the shape parameter λ, which is unknown. Therefore we resort to a parametric bootstrap procedure in order to obtain the critical point p_α of the test as follows:

- 1. Conditionally on the observed value of Y_j, $j = 1, 2, \ldots, n$, compute the estimates $(\hat{\delta}_n, \hat{c}_n, \hat{\lambda}_n)$ and then the observations $\hat{X}_j = (Y_j - \hat{\delta}_n)/\hat{c}_n$, $j = 1, 2, \ldots, n$.

- 2.a. Calculate the value of the test statistic, say $\hat{T}$, based on $\hat{X}_j$ and $\hat{\lambda}_n$.
- 2.b.1. Generate a bootstrap sample Y_j^*, $j = 1, 2, \ldots, n$, from $SNVG(0, 1, \hat{\lambda}_n)$.
- 2.b.2. On the basis of Y_j^*, $j = 1, 2, \ldots, n$, compute the estimates $(\hat{\delta}_n^*, \hat{c}_n^*, \hat{\lambda}_n^*)$ and then the observations $\hat{X}_j^* = (Y_j^* - \hat{\delta}_n^*)/\hat{c}_n^*$, $j = 1, 2, \ldots, n$.
- 2.b.3. Calculate the value of the test statistic, say $\hat{T}^*$, based on $\hat{X}_j^*$ and $\hat{\lambda}_n^*$.
- 3. Repeat steps 2.b.1 - 2.b.3., and calculate M values of $\hat{T}^*$, say $\hat{T}_j^*$, $j = 1, 2, \ldots, M$.
- 4. Obtain p_α as $\tilde{T}^*_{(M-\alpha M)}$, where $\tilde{T}^*_{(j)}$, $j = 1, 2, \ldots, M$ denotes the ordered $\tilde{T}_j^*$ - value.

Acknowledgements: The author wishes to sincerely thank S. Meintanis and D. Karlis for their strong support and guidance during this work.

5. Bibliography

[1] N. Gürtler, and N. Henze, "Goodness–of–fit tests for the Cauchy distribution based on the empirical characteristic function", *Ann. Instit. Statist. Math.* **52**, 267–286 (2000).

[2] B. Klar, and S. Meintanis, "Tests for normal mixtures based on the empirical characteristic function", *Comput. Statist. Dat. Anal.* **49**, 227–242 (2005).

[3] I. A. Koutrouvelis, and S. G. Meintanis, "Testing for stability based on the empirical characteristic function with applications to financial data", *J. Statist. Comput. Simulation* **64**, 275–300 (1999).

[4] D. Karlis, "An EM Type Algorithm for ML estimation for the Normal - Inverse Gaussian Distribution", *Statistics and Probability*, **57**, 43-52 (2002)