Goodness of Fit Test for α-stable Distribution Based on Quantile Conditional Variance Statistics

Srinivasan Sathiamurthy

Authors: Pitera, Chechkin, Wylomanska

December 1, 2022

Stable Distributions

Definition 1: Stable Distribution

A random variable X has a stable distribution if for any $A, B \in \mathbb{R}^+, \exists C \in \mathbb{R}^+, \exists D \in \mathbb{R}$ such that

$$AX_1 + BX_2 \stackrel{d}{=} CX + D$$

Where X_1, X_2 are independent copies of X. X is strictly stable if $D = 0$.

Theorem 2: α-stable Distribution

For any stable random variable X, there is a number $\alpha \in (0, 2]$ such that the number C satisfies:

$$C^\alpha = A^\alpha + B^\alpha$$

Feller (1971) proved $C_n = n^{\frac{1}{\alpha}}$ for some $0 < \alpha \leq 2$

α-Stable Distributions

Definition 3: Characteristic Function of α-stable Distributions

We say X follows the α-stable distribution if its characteristic function is given by:

$$\mathbb{E}\left[e^{itX}\right] = \begin{cases} \exp\left\{-c^{\alpha}|t|^{\alpha}\left(1 - i\beta \operatorname{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mu t\right\} & \alpha \neq 1 \\ \exp\left\{-c|t|\left(1 + i\beta\frac{2}{\pi}\operatorname{sgn}(t)\ln(|t|)\right) + i\mu t\right\} & \alpha = 1 \end{cases}$$

where $\alpha \in (0,2], \beta \in [-1,1], c > 0, \mu \in \mathbb{R}$. For brevity, we write $X \sim S(\alpha, c, \beta, \mu)$. If $\beta = 0$, we say $X \sim S\alpha S(\alpha, c, \mu)$

Remark 1: This is equivalent to Definition 1.

Remark 2: When $\alpha < 2$, the second moment does not exist.

Remark 3: If $\alpha < 1$, then the first moment does not exist.

Special Cases

1. Gaussian Distribution: $S(2, c, 0, \mu) = N(\mu, 2c^2)$, with density

$$\frac{1}{2c\sqrt{\pi}} \cdot e^{-\frac{(x-\mu)^2}{4c^2}}$$

2. Cauchy Distribution: $S(1, c, 0, \mu)$, with density:

$$\frac{c}{\pi\left((x-\mu)^2 + c^2\right)}$$

3. Levy distribution: $S\left(\frac{1}{2}, c, 1, \mu\right)$, with density concentrated on (μ, ∞):

$$\left(\frac{c}{2\pi}\right)^{\frac{1}{2}} \frac{1}{(x-\mu)^{\frac{3}{2}}} \exp\left\{-\frac{c}{2(x-\mu)}\right\}$$

4. $S(\alpha, 0, 0, \mu)$ is degenerate at μ

Properties of Stable Random Variables: Shift and Scale

Property 1: μ is a shift parameter

If $X \sim S(\alpha, c, \beta, \mu)$, and $a \in \mathbb{R}$, then $X + a \sim S(\alpha, c, \beta, \mu + a)$

Property 2: c is a scale parameter

If $X \sim S(\alpha, c, \beta, \mu)$, and $a \in \mathbb{R}, a \neq 0$, then:

$$\alpha \neq 1 \Rightarrow aX \sim S(\alpha, |a|c, \operatorname{sgn}(a)\beta, a\mu)$$

$$\alpha = 1 \Rightarrow aX \sim S(1, |a|c, \operatorname{sgn}(a)\beta, a\mu - \frac{2}{\pi}a(\ln|a|)c\beta)$$

c isn't really a scale parameter when $\alpha = 1, \beta \neq 0$ due to property 2. But when $\mu = 0$ in that case as well and $0 < \alpha < 2$:

$$X \sim S(\alpha, c, \beta, 0) \iff -X \sim S(\alpha, c, -\beta, 0)$$

Properties of Stable Random Variables: Skewness and Stability

Property 3: β is a skewness parameter

$X \sim S(\alpha, c, \beta, \mu)$ is symmetric about μ iff $\beta = 0$. (Follows from requiring the characteristic function to be real)

Property 4: Stabilty and Symmetry for $\alpha \neq 1$

If $\alpha \neq 1$, and $X \sim S(\alpha, c, \beta, \mu)$, then $X - \mu$ is strictly stable.

Property 5: Stabilty and Symmetry for $\alpha = 1$

$X \sim S(1, c, \beta, \mu)$ is strictly stable iff $\beta = 0$

$S(\alpha, c, \beta, 0)$ is said to be skewed to the right if $\beta > 0$ and to the left if $\beta < 0$. It is totally skewed to the right if $\beta = 1$ and totally skewed to the left if $\beta = -1$

Simulation of α-stable Distributions

Proposition 4: Simulation for $S(\alpha, 1, 0, 0)$

Let γ be uniform on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and let W be exponential with mean 1. Assume γ, W are independent. Then:

$$X = \frac{\sin(\alpha\gamma)}{(\cos\gamma)^{\frac{1}{\alpha}}} \left(\frac{\cos((1-\alpha)\gamma)}{W}\right)^{\frac{1-\alpha}{\alpha}} \sim S(\alpha, 1, 0, 0)$$

Combining proposition 4 with property 1 and 2 allows one to simulate $S(\alpha, c, 0, \mu)$, such as via Monte Carlo simulations, which is used extensively in the literature.

Stable Densities for Varying Alphas with Beta = 0

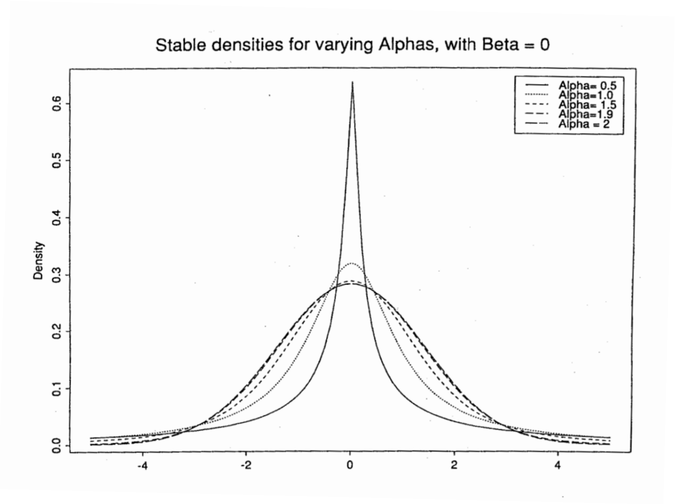

Figure: The paper only considers cases when $\beta = 0$

Quantile Conditional Variance

Definition 5: Quantile Conditional Variance

For some R.V X, and any $0 \leq a < b \leq 1$, define $A_x(a,b) := \{X \in (F_X^{-1}(a), F_X^{-1}(b))\}$. The QCV of X is:

$$\sigma_X^2(a,b) := \operatorname{Var}[X | A_X(a,b)]$$

Theorem 6: QCVs can act as law classifiers

Let X, Y be absolutely continuous R.Vs such that $\sigma_X^2(a,b) = \sigma_Y^2(a,b)$ for $0 \leq a < b \leq 1$. Then there exists $\mu \in \mathbb{R}$ such that

$$F_X(t) = F_{Y+\mu}(t), \forall t \in \mathbb{R}$$

Remark: The limit values $a = 0, b = 1$ are not required for full classification due to limit arguments.

Sample QCV

Definition 7: Sample QCV

Given i.i.d sample $X_1, X_2, \ldots, X_n$, and quantile values $0 < a < b < 1$, the sample estimator of $\sigma_X^2(a, b)$ is given by:

$$\hat{\sigma}_X^2(a, b) := \frac{1}{\lfloor nb \rfloor - \lfloor na \rfloor} \sum_{i=\lfloor na \rfloor + 1}^{\lfloor nb \rfloor} \left(X_{(i)} - \hat{\mu}_X(a, b) \right)^2$$

Where $X_{(i)}$ is the ith order statistic of the sample and $\hat{\mu}_X(a, b)$ is the conditional sample mean (average $X_{(i)}$ over $\lfloor na \rfloor + 1 \leq i \leq \lfloor nb \rfloor$)

QCV For α-stable Distributed R.Vs

Proposition 8: QCV estimator is consistent and asymptotically normal

Let $X \sim S(\alpha, c, \beta, \mu)$. Then for any $0 < a < b < 1$, when $n \to \infty$:

$$\hat{\sigma}_X^2(a,b) \stackrel{\mathbb{P}}{=} \sigma_X^2(a,b)$$

And also when $n \to \infty$:

$$\sqrt{n}\left(\hat{\sigma}_X^2(a,b) - \sigma_X^2(a,b)\right) \stackrel{d}{=} N(0, c\tau)$$

where $\tau > 0$ is a fixed constant depending on α, β, a, b.

Remark 1: Typically, the constant τ is approximated using Monte Carlo simulations.

Remark 2: Any linear combination of QCV sample estimators is also asymptotically normal.

Location and Scale Testing for $S(\alpha, c, 0, \mu)$ Distributions

- Matsui and Takemura propose testing the null hypothesis that a random sample $x_1, \ldots, x_n$ from an unknown distribution F belongs to the family of non-skewed stable distributions $S(\alpha, c, 0, \mu)$.
- They find the difference between the characteristic function of an α-stable distribution with α estimated from the data, and the empirical characteristic function after standardizing $(\mu, c) = (0, 1)$:

$$\Phi_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp\left(-i \cdot t \cdot \frac{x_j - \hat{\mu}}{\hat{c}}\right)$$

- They use a weighted L^2-distance as their test statistic:

$$D_{n,\kappa} := n \int_{-\infty}^{\infty} \left|\Phi_n(t) - e^{-|t|^{\hat{\alpha}}}\right|^2 w(t)dt, w(t) = e^{-\kappa|t|}, \kappa > 0$$

Stability Index Testing for $S(\alpha, c, 0, \mu)$ Distributions

- This paper concerns with α since that is linked to tail behavior: they compare tail QCVs with central region QCVs to assess how heavy are the tails
- Generic family of test statistics that they propose for goodness-of-fit testing for any distribution:

$$N := \sqrt{n}\frac{d_1\hat{\sigma}^2_X(a_1, a_2) + d_2\hat{\sigma}^2_X(a_2, a_3) + d_3\hat{\sigma}^2_X(a_3, a_4)}{\hat{\sigma}^2_X(a_1, a_4)}$$

where $x = (x_1, \ldots, x_n)$ is a sample from the R.V X's distribution, $d_1, d_2, d_3 \in \mathbb{R}$ are fixed weight parameters, and $0 < a_1 < a_2 < a_3 < a_4 < 1$ are the quantile split parameters.

Stability Index Testing for $S(\alpha, c, 0, \mu)$ Distributions

- $\hat{\sigma}^2_X(a_1, a_4)$ on the denominator makes N invariant to affine transformations of X, so (μ, c) don't matter anymore.
- Since X is symmetric ($\beta = 0$), they set $a_4 = 1 - a_1$, $a_3 = 1 - a_2$, $d_1 = d_3$.
- For a given a_1, a_2, they fix d_1, d_2 such that N is normalized when $\alpha = 2$.
- They consider three quantile splits:
  1. 5/20/50/20/5 compares the central region QCV with the trimmed tail QCVs. The top and bottom 5% ensure finiteness of tail QCVs without inducing severe volatility.
  2. 0.5/24.5/50/24.5/0.5 Increases sensitivity of tail QCVs, allowing more accurate α testing when the underlying sample size is large.
  3. 0.5/3.5/92/3.5/0.5 Detects minor α changes for large sample sizes.

N-statistic Analysis

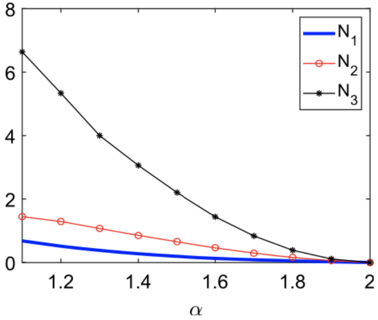

Figure: Values of N-statistic for $S(\alpha, 1, 0, 0)$ for the three quantile specifications.

N_1, N_2, N_3 correspond to the first, second, and third quantile splits, respectively.

N-statistic Analysis

- N_3 captures the α better than N_2, which does better than N_1. So given enough data and depending on how much you want to discriminate α parameters close to 2, pick a higher N.
- Given a specific sample at hand, don't tune a_i, d_i since that would reduce the statistical power of the test.

Power Simulation Study of N-statistic Setup

- They perform the analysis under H_0 stating that the sample comes from $S(1.5, 1, 0, 0)$. (Two-sided)
- Values of N_1, N_2, N_3 statistics under H_0 are calculated based on 100,000 Monte Carlo simulations for $n \in \{20, 50, 100, 500, 1000, 2000\}$
- They check the powers of the tests against H_1 hypotheses stating that the random sample comes from $S(\alpha_1, 1, 0, 0)$ for $\alpha_1 \in \{1.1, 1.2, \ldots, 2.0\}$ using $10,000$ Monte Carlo simulations for each α_1.
- QCV tests are compared against Kolmogorov-Smirnov cumulative probability test (KS), Kuiper test, Watson test, Cramer-von Mises test (CvM) and the Anderson-Darling test (AD).

Power Simulation Study of N-statistic Results

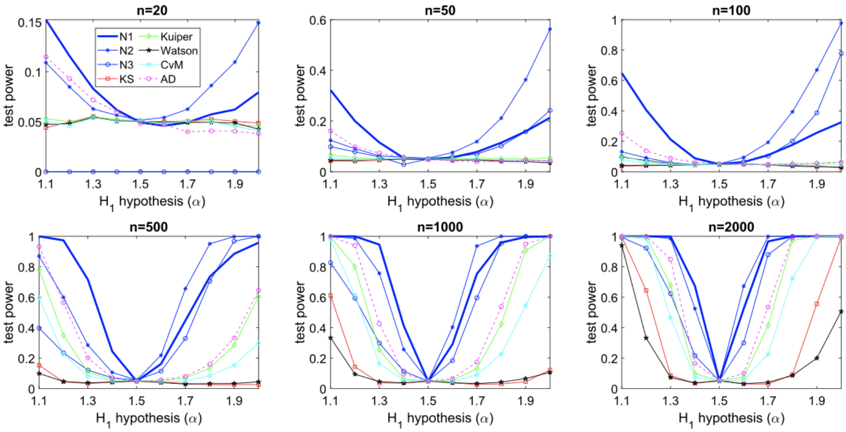

Figure: Test power for significance level 5%. Results for 1% were similar.

Power Simulation Study of N-statistic Results

- QCV statistics outperform benchmark frameworks
- QCV statistics allow more efficient local density discrimination
- N-statistic is tailored to measure fat-tail behavior - efficient when assessing the value of α
- Apart from α close to 2, the power of N_1 has the best fit among all the statistics
- Smaller α makes it harder to estimate the extreme 0.5% quantile needed for N_2, N_3
- For $\alpha \geq 1.5$, N_2 outperforms N_1.
- N_3 starts to outperform N_2 for $\alpha \geq 1.92$ for two-sided tests at 1% and 5% significance

Empirical Case Study of Market Stock Returns

- Consider a simpler N-statistic with a 20-60-20 quantile split
- 1 year of return rates (GOOGL, AAPL, SNP500, DAX) - unaligned
- Tests normality vs not normal due to tail

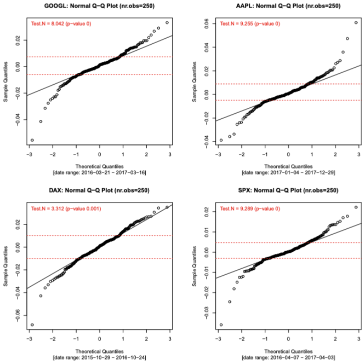

Figure: Normal fit is good for middle 60% of observations, bad for the tails

Simple N-statistic Rejection Analysis Setup

- Take returns of all stocks listed in SNP500 on 6/16/2018 that have full historical data from 1/2000 to 5/2018
- Split returns into disjoint sets of length $n \in \{50, 100, 250\}$
- For each subset, compare the value of the N-statistic with the corresponding empirical quantiles (computed earlier)
- Perform right-sided statistical test and reject normality if computed N-statistic is greater than the computed cdf plot for significance levels $\alpha \in \{1\%, 2.5\%, 5\%\}$

Simple N-statistic Rejection Analysis Table

- Statistic T: Proportion of data on which the normality assumption was rejected at the significance level (α)
- Statistic U: Proportion of data on which normality assumption was rejected at a given significance level (α) only by the considered test (among Jarque-Bera test, Anderson-Darling test, Shapiro-Wilk test, N-test)
- *nr*-runs indicates how many subsamples were created for a given *n*

Simple N-statistic Rejection Analysis Table

Desc	nr runs	α	n	rejects	JB	AD	SW	N
T U	35052	1.0%	50	31.5%	**25.9%** 1.9%	17.3% 0.6%	23.2% 0.3%	**25.9%** **3.1%**
T U	35052	2.5%	50	39.6%	32.4% 2.4%	22.9% 0.9%	28.3% 0.3%	**32.5%** **3.9%**
T U	35052	5.0%	50	47.9%	38.6% 2.4%	29.1% 1.3%	33.7% 0.4%	**39.6%** **5.1%**
T U	17526	1.0%	100	52.8%	45.2% 2.2%	31.8% 0.6%	41.3% 0.3%	**46.1%** **4.4%**
T U	17526	2.5%	100	61.3%	52.9% 2.2%	38.8% 0.7%	47.6% 0.2%	**54.3%** **5.1%**
T U	17526	5.0%	100	68.4%	59.7% 2.2%	45.7% 0.8%	53.4% 0.2%	**61.3%** **5.3%**
T U	6858	1.0%	250	88.5%	82.1% 1.0%	71.2% 0.4%	79.3% 0.1%	**85.4%** **3.8%**
T U	6858	2.5%	250	91.8%	86.8% 0.7%	77.7% 0.2%	83.7% 0.1%	**89.4%** **3.0%**
T U	6858	5.0%	250	93.9%	89.7% 0.5%	82.4% 0.2%	86.9% 0.0%	**92.0%** **2.5%**

Figure: *N*-statistic dominates

References

1. Stable non-gaussian random processes, *Samorodnitsky & Taqqu*
2. Goodness-of-fit test for α-stable distribution on the quantile conditional variance statistics, *Pitera, Chechkin & Wylomanska*
3. Goodness-of-fit tests for symmetric stable distributions - empirical characteristic function approach, *Matsui & Takemura*
4. New fat-tail normality test based on conditional second moments with applications to finance, *Jelito & Pitera*