



New fat-tail normality test based on conditional second moments with applications to finance

Damian Jelito¹ · Marcin Pitera¹

Received: 5 March 2019 / Revised: 30 December 2019 / Published online: 21 May 2020
© The Author(s) 2020

Abstract

In this paper we introduce an efficient fat-tail measurement framework that is based on the conditional second moments. We construct a goodness-of-fit statistic that has a direct interpretation and can be used to assess the impact of fat-tails on central data conditional dispersion. Next, we show how to use this framework to construct a powerful normality test. In particular, we compare our methodology to various popular normality tests, including the Jarque–Bera test that is based on third and fourth moments, and show that in many cases our framework outperforms all others, both on simulated and market stock data. Finally, we derive asymptotic distributions for conditional mean and variance estimators, and use this to show asymptotic normality of the proposed test statistic.

Keywords 20-60-20 rule · Normality test · Fat-tail · Heavy-tail · Non-normality · Stock returns

Mathematics Subject Classification 62F03 · 62F05 · 62P05 · 62P20 · 91G70

1 Introduction

It has been recently shown in Jaworski and Pitera (2016) that for a normal random variable and a unique ratio close to 20/60/20 the conditional dispersion in the tail sets

Part of the work of the second author was supported by the National Science Centre, Poland, via project 2016/23/B/ST1/00479.

The authors would like to thank the anonymous referees for their helpful comments and suggestions which improved greatly the final manuscript.

✉ Damian Jelito
damian.jelito@im.uj.edu.pl

Marcin Pitera
marcin.pitera@im.uj.edu.pl

¹ Institute of Mathematics, Jagiellonian University, Łojasiewicza 6, 30-348 Cracow, Poland

is the same as in the central set. In other words, if we split big normal sample into three sets—one corresponding to the worst 20% outcomes, one corresponding to the middle 60% outcomes, and one corresponding to the best 20% outcomes—then the conditional variance on those subsets is approximately the same.

In this paper we show that this property could be used to construct an efficient goodness-of-fit testing framework that has a direct (financial) interpretation. The impact of tail dispersion on central dispersion is a natural measure of tail heaviness and can serve as an alternative to other methods which are typically based on tail limit analysis or higher order moments; see Alexander (2009) and Jarque and Bera (1980). In particular, in contrast to the Jarque–Bera normality test that is based on third and fourth moments, our test relies on the conditional second moments which are often easier to estimate.

Testing for normality has a long history and many remarkable methods have been developed. This includes general distribution-fit frameworks like Anderson–Darling test based on the distance between theoretical and empirical distribution function (Anderson and Darling 1954) or Shapiro–Wilk test relying on the regression coefficient (Wilk and Shapiro 1965); see Madansky (2012), Henze (2002) or Thode (2002) for a comprehensive overview of normality testing procedures.

Most empirical studies suggest that the normality tests should be chosen carefully as their statistical power varies depending on the context; see e.g. Thadewald and Büning (2007), Romão et al. (2010) or Brockwell and Davis (2016). This is why the existing procedures are constantly refined and new ones are being developed; for example, a recent revision of the Jarque–Bera testing framework based on the second-order analogue of skewness and kurtosis could be found in Desgagné and Lafaye de Micheaux (2018).

The approach presented in this paper draws attention to interesting and previously not exploited aspect of normal distributions that could be used for efficient normality testing. In particular, we show that our approach outperforms and/or complements multiple benchmark normality testing frameworks when popular (financial) alternative distributions, such as Student's *t* or logistic, are considered. More explicitly, we show that our test usually has the best power if one expects that a sample comes from a symmetric distribution which have heavier (or lighter) tails in comparison with the normal distribution. We illustrate that with a financial market data example; see Sect. 6 for details.

Finally, it is worth mentioning that the 20/60/20 division leads to a very accurate data clustering when it comes to the tail assessment performed in reference to the central set normality assumption. Consequently, our method could be embedded into data analytics frameworks based on cluster analysis, help to refine data mining techniques, etc; see Romesburg (2004), Kaufman and Rousseeuw (2009) and Hair et al. (2013) for an overview. In fact, the good performance of our test statistic on market data could be linked to a popular financial *stylised fact* saying that typical financial asset returns can be seen as normal, but the extreme returns are more frequent and with greater magnitude than the ones resulting from the normal fit; see Cont (2001) and Sheikh and Qiao (2010) for details.

This paper is organised as follows: In Sect. 2 we briefly recall the concept of the 20-60-20 Rule, while in Sect. 3 we outline the construction of the test statistic and discuss

its basic properties. Section 4 provides a high-level discussion about the test power, and Sect. 5 discusses in details mathematical background including derivation of the asymptotic distribution of the proposed test statistic. Next, in Sect. 6, we present a simple market data case-study and discuss application of our framework in the financial context. We conclude in Sect. 7. For brevity, we moved the closed-form formula for the normalising constant introduced in Sect. 3 to Appendix A.

2 The 20-60-20 rule for the univariate normal distribution

Let us assume that X is a normally distributed random variable. We define left, middle, and right partitioning sets of X by

$$L := \left(-\infty, F_X^{-1}(0.2)\right], M := \left(F_X^{-1}(0.2), F_X^{-1}(0.8)\right), R := \left[F_X^{-1}(0.8), +\infty\right), \tag{1}$$

where $F_X^{-1}(\alpha)$ is the α -quantile of X . It has been shown in Jaworski and Pitera (2016) that

$$\sigma_L^2 = \sigma_M^2 = \sigma_R^2, \tag{2}$$

for this unique 20/60/20 ratio, where σ_A^2 denotes the conditional variance of X on set A .¹

This specific division together with the associated set of equalities in (2) create a dispersion balance for the conditioned populations. This property might be linked to the statistical phenomenon known as the *20-60-20 Rule*: a principle that is widely recognised by the practitioners and used e.g. for efficient management or clustering. In fact, a similar statement is true in the multivariate case: the conditional covariance matrices of multivariate normal vector are equal to each other, when the conditioning is based on the values of any linear combination of the margins, and 20/60/20 ratio is maintained. For more details, see Jaworski and Pitera (2016) and references therein.

3 Test statistic

Let us assume we have a sample from X at hand. Then, based on (2), we define a test statistic

$$N := \frac{1}{\rho} \left(\frac{\hat{\sigma}_L^2 - \hat{\sigma}_M^2}{\hat{\sigma}^2} + \frac{\hat{\sigma}_R^2 - \hat{\sigma}_M^2}{\hat{\sigma}^2} \right) \sqrt{n}, \tag{3}$$

where $\hat{\sigma}^2$ is the sample variance, $\hat{\sigma}_A^2$ is the conditional sample variance on set A (where the conditioning is based on empirical quantiles), n is the sample size, and $\rho \approx 1.8$ is a fixed normalising constant; see Fig. 1 for the **R** implementation code. We refer to Sect. 5 for more details including rigorous definitions of conditional variance $\hat{\sigma}_A^2$, constant ρ , etc.

¹ In fact, this equality is true for the ratio very close to 20/60/20, i.e. for upper and lower quantiles equal to approximately 0.198. For transparency, we have decided to use the rounded numbers here; see Sect. 5 for details.

```

1 Test.N <- function(x){
  n <- length(x)
3  q1 <- quantile(x,0.2)
  q2 <- quantile(x,0.8)
5  low <- x[x <= q1]
  med <- x[x > q1 & x < q2]
7  high <- x[x >= q2]
  N <- var(low)+var(high)-2*var(med)
9  N <- N * sqrt(n)/(var(x)*1.8)
  return(N)}

```

Fig. 1 Simplified **R** source code that was used to create a function that computes the test statistic N given input sample x

It is not hard to see that under the normality assumption N is a pivotal quantity. Furthermore, in Sect. 5 we show that the distribution of N is asymptotically normal; see Theorem 1 therein. In Fig. 2, we illustrate this by computing the Monte Carlo density of N under the normality assumption for samples of size 50, 100, and 250.

Test statistic N has a clear interpretation: the difference between tail and central conditional variances could be seen as a measure of tail fatness, i.e. the bigger the value of N , the fatter the tails.

To illustrate this we compute the values of N for bigger sample size, $n = 500$, for three different fat-tail distributions and three different slim-tail distributions. For fat-tail comparison we picked logistic, Student's t with five degrees of freedom, and Laplace distributions, while for slim-tail comparison we considered generalised normal distribution with shape parameter $s \in \{2.5, 3, 5\}$; the (standardised) generalised normal density for $s \in \mathbb{R}_+$ is given by

$$f(x|s) := \frac{s}{2\Gamma(1/s)} \exp(-|x|^s), \quad x \in \mathbb{R}; \quad (4)$$

we refer to Nadarajah (2005) or Tumlinson et al. (2016) for more details. The results presented in Fig. 3 confirm that the behaviour of N is as expected.

Based on statistic N values, one can construct a one-sided or two-sided statistical test with normality ($N = 0$) as a null hypothesis. For brevity, we refer to such test as N normality test or simply N test.

4 Power of the test

In this section we check the power of the proposed N test in a controlled environment. We focus on symmetric distributional alternatives (used e.g. in finance) when one wants to abandon the normality assumption due to fat-tail or slim-tail phenomena. Namely, we consider the Cauchy distribution, the Logistic distribution, the Laplace distribution, the Student's t distribution with $v \in \{2, 5, 10, 20, 30\}$ degrees of freedom parameter, and the generalised normal distribution (GN) with the shape parameter $s \in \{1.5, 2.5, 3, 5, 10\}$. Note that GN with $s = 1$ and $s = 2$ correspond to Laplace and

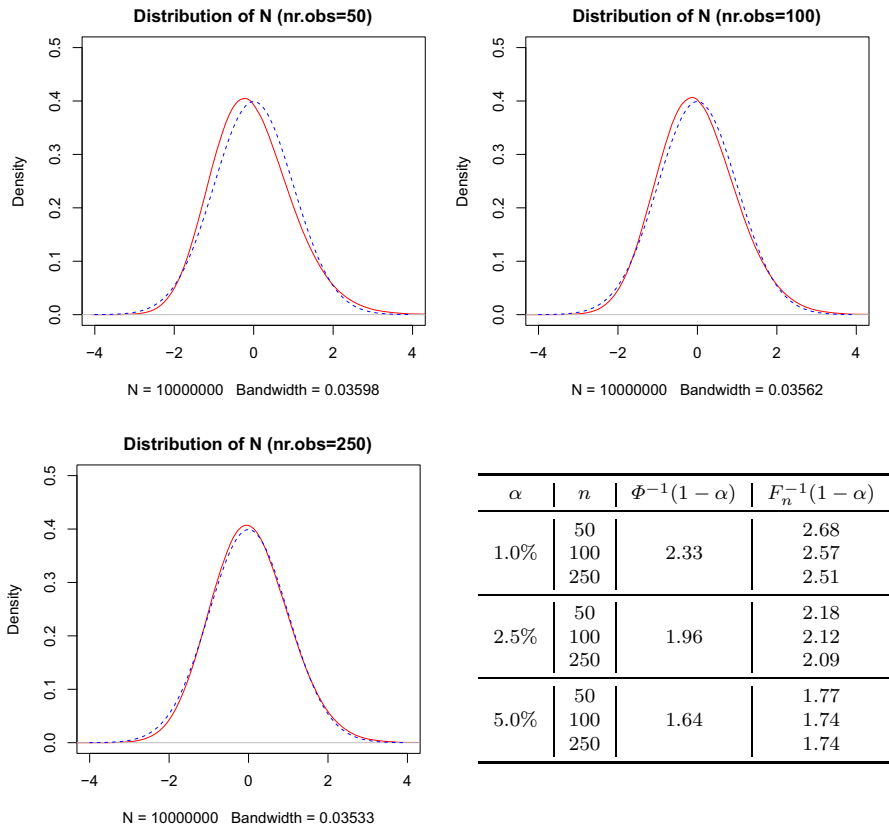


Fig. 2 The distribution of N under the normality assumption for $n = 50, 100, 250$, for strong Monte Carlo sample of size 10,000,000. The obtained empirical density (solid curve) is very close to standard normal density (dashed curve); the table compares selected empirical quantiles with theoretical normal quantiles

normal distribution, respectively; see (4) for the GN density definition. In all cases, the location parameter is set to 0 and the scale parameter is set to 1.

For completeness, we compare test N with well-established benchmark normality tests: Jarque–Bera test, Anderson–Darling test, and Shapiro–Wilk test. It should be noted that in contrast to these frameworks, statistic N allows one to consider a specific heavy-tail (or slim-tail) alternative, i.e. positive (or negative) values of N point out to heavy-tail (or slim-tail) alternative hypothesis. Consequently, we decided to construct right-sided (or left-sided) critical region for the fat-tailed (or slim-tailed) distributions. Nevertheless, for completeness, we include the results for two-sided critical regions in all cases.

For all alternative distribution choices we consider four different sample sizes, i.e. $n = 20, 50, 100, 250$. For each n , we simulate 2,000,000 strong Monte Carlo sample, and check for what proportion of simulations the tests reject normality at significance level $\alpha = 5\%$.

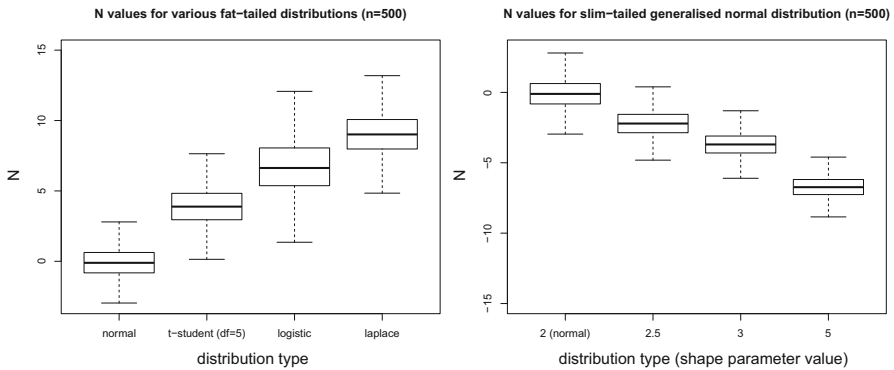


Fig. 3 Boxplots of N for samples from different distributions. The results for three fat-tailed distributions (Logistic, t-Student with $\nu = 5$, and Laplace) are presented in the left plot. The results for three slim-tailed distributions (generalised normal with $s \in \{2.5, 3, 5\}$) are presented in the right plot. For transparency, we rescaled all distributions to have zero mean and unit variance, and added the results for normal distribution. The boxplots are based on strong Monte Carlo sample of size 10,000, where each simulation is of size $n = 500$

All computations are performed in **R 3.5.2**. For benchmark normality testing we use multiple add-on **R** packages including *gnorm* (for GN simulation), *stats* (for Shapiro–Wilk test), *nortest* (Anderson–Darling test), and *tseries* (Jarque–Bera test). For better comparability, we used test-wise simulated rejection thresholds instead of theoretical p -values returned by **R** functions; for computations, we used big strong Monte Carlo sample of size 10,000,000. In particular, note that while Jarque–Bera test has asymptotic χ^2 distribution (with 2 degrees of freedom) under normality, this approximation may be inaccurate for small samples, and lead to non-meaningful (non-adjusted) p -values.

It should be noted that our framework specification is consistent with the one presented in Desgagné and Lafaye de Micheaux (2018), where a comprehensive normality tests comparison is made. In particular, results presented in Appendix C therein are perfectly consistent with results presented here (for all benchmark tests).²

For transparency, we have decided to consider fat-tail and slim-tail case separately. The results for fat-tailed distributions are presented in Table 1. The best performance of the right-sided test N could be observed for almost all considered distributions. The fatter the tails the bigger the absolute difference between N test power and JB test power, which could be considered as a second best choice. To check whether test statistic N brings some novel results, we decided to check the proportion of simulations on which the normality assumption was rejected uniquely by N among all considered tests; for comparison purposes, we checked the same for all other tests. The results for three selected distributions are presented in Table 2. It can be observed, that the unique rejection proportion for test N is the highest among all tests and point out to the fact that statistic N is taking into account sample properties that are not exploited

² Please note that results presented for Asymmetric Power Distribution (APD) with symmetry parameter $\alpha = 0.5$ correspond to results presented for Generalised Normal (GN) distribution; we refer to Section 2 in Desgagné and Lafaye de Micheaux (2018) for details.

by other tests.³ Finally, it should be noted that the performance of two-sided N test is also quite good: the test outperforms both AD and SW in most of the considered cases.

The results for slim-tailed distributions are presented in Table 3. Both left-sided and two-sided test based on N statistic substantially outperform all other tests on all datasets. This suggest that the proposed framework is also suitable when assessing slim-tailed distributions.

5 Mathematical framework and asymptotic results

In this section, we provide the explicit formulas for the conditional variance estimators, study their asymptotic behaviour, and show that N is asymptotically normal.

First, we introduce the basic notation and provide more explicit formulas for sets L , M , and R that were given in Sect. 2; see (1).

We assume that $X \sim \mathcal{N}(\mu, \sigma)$ for mean parameter μ and standard deviation parameter σ . We use F_X to denote the distribution of X , Φ to denote the standard normal distribution, and ϕ to denote the standard normal density. Following the usual convention, for any $n \in \mathbb{N}$, we use (X_1, \dots, X_n) to denote the random sample from X and for $i = 1, \dots, n$, we use $X_{(i)}$ to denote the sample i th order statistic.

For fixed partition parameters $\alpha, \beta \in \mathbb{R}$, where $0 \leq \alpha < \beta \leq 1$, we define the conditioning set

$$A[\alpha, \beta] := \{x \in \mathbb{R} : F_X^{-1}(\alpha) < x \leq F_X^{-1}(\beta)\}.$$

For brevity and with slight abuse of notation, we often write A instead of $A[\alpha, \beta]$. Then, the explicit formulas for sets L , M , and R given in (1) are

$$L := A[0, \tilde{q}], \quad M := A[\tilde{q}, 1 - \tilde{q}], \quad R := A[1 - \tilde{q}, 1], \tag{5}$$

where $\tilde{q} := \Phi(x)$, and x is the unique negative solution of the equation

$$-x\Phi(x) - \phi(x)(1 - 2\Phi(x)) = 0. \tag{6}$$

The approximate value of \tilde{q} is 0.19809; we refer to (Jaworski and Pitera 2016, Lemma 3.3) for details. Note that (6) could be seen as a specific form of differential equation $-xy - y'(1 - 2y) = 0$, where $y(x) := \Phi(x)$; this could be used to determine similar ratios for other distributions.

Next, we give the exact definition of the conditional sample variance. For a fixed set A , where $A = A[\alpha, \beta]$, the conditional variance estimator on the set A is given by

$$\hat{\sigma}_A^2 := \frac{1}{[n\beta] - [n\alpha]} \sum_{i=[n\alpha]+1}^{[n\beta]} (X_{(i)} - \bar{X}_A)^2, \tag{7}$$

³ The results for other distributions are consistent with those presented in Table 2 and could be obtained from authors upon request.

Table 1 The table contains **test power** for various fat-tailed alternatives at significance level $\alpha = 5\%$

Distr	n	JB (%)	AD (%)	SW (%)	N (two-sided) (%)	N (right-sided) (%)
Cauchy	20	86.2	88.0	86.6	85.5	89.0
	50	99.5	99.7	99.6	99.7	99.8
	100	100.0	100.0	100.0	100.0	100.0
	250	100.0	100.0	100.0	100.0	100.0
Logistic	20	14.9	10.6	11.7	11.1	15.6
	50	25.9	15.9	19.6	21.2	28.9
	100	39.4	23.9	30.5	35.8	45.6
	250	66.7	46.7	57.0	67.8	76.4
Student's t (2)	20	56.7	52.9	52.9	52.9	60.1
	50	88.2	85.8	86.3	89.2	92.3
	100	98.8	98.4	98.6	99.2	99.6
	250	100.0	100.0	100.0	100.0	100.0
Student's t (5)	20	22.9	17.1	18.6	18.1	23.9
	50	43.0	30.2	35.5	38.4	46.8
	100	64.6	48.1	56.4	62.5	70.5
	250	92.0	81.8	88.2	92.8	95.4
Student's t (10)	20	12.3	8.8	9.8	9.2	12.5
	50	20.6	12.0	15.5	15.7	21.6
	100	30.7	16.2	23.0	24.9	33.1
	250	52.3	28.5	41.7	47.8	57.9
Student's t (20)	20	8.1	6.4	6.8	6.5	8.2
	50	11.4	7.2	8.7	8.4	11.5
	100	15.1	8.0	11.0	10.7	15.5
	250	22.9	10.3	16.2	17.2	24.8
Student's t (30)	20	6.9	5.8	6.1	5.9	7.0
	50	8.9	6.2	7.1	6.8	8.9
	100	10.9	6.5	8.3	7.9	11.1
	250	15.0	7.4	10.7	10.7	16.0
Laplace/GN (1)	20	30.2	27.2	26.1	26.4	35.4
	50	55.6	54.5	52.1	59.3	69.1
	100	79.9	82.7	79.7	87.3	92.0
	250	98.7	99.6	99.2	99.8	99.9
GN (1.5)	20	12.1	9.1	9.5	9.1	13.4
	50	19.0	13.2	14.4	16.1	23.7
	100	27.9	19.8	21.5	27.3	37.5
	250	48.6	40.2	41.4	55.8	66.9

Distr refers to the distribution used for the alternative hypothesis and *n* indicates sample size. The acronyms JB, AD, SW, and N refer to Jarque–Bera, Anderson–Darling, Shapiro–Wilk, and *N* test, respectively. Best performance is marked in bold.

Table 2 The table contains the **unique rejections ratios** for three selected fat-tailed alternative distribution

Distr	n	JB (%)	AD (%)	SW (%)	N (right-sided) (%)
Logistic	20	2.2	1.0	0.4	4.3
	50	2.7	1.2	0.4	6.9
	100	2.7	1.0	0.3	9.3
	250	1.6	0.5	0.1	9.7
Student's t (20)	20	1.6	0.9	0.4	2.8
	50	2.1	1.3	0.5	3.6
	100	2.5	1.4	0.5	4.7
	250	3.0	1.3	0.5	7.0
Laplace/GN (1)	20	2.1	1.6	0.2	7.6
	50	1.2	1.6	0.1	9.2
	100	0.3	0.7	0.0	5.1
	250	0.0	0.0	0.0	0.2

For each test, a proportion of simulated samples on which normality hypothesis was uniquely rejected by the test (among all four considered tests) is computed. The remaining notation is aligned with Table 1 Best performance is marked in bold

Table 3 The table contains **test power** for various slim-tailed alternatives at significance level $\alpha = 5\%$

Distr	n	JB (%)	AD (%)	SW (%)	N (two-sided) (%)	N (left-sided) (%)
GN (2.5)	20	2.4	4.5	4.1	5.4	8.6
	50	1.3	5.5	4.6	8.0	13.6
	100	0.9	7.4	6.2	13.0	21.1
	250	4.3	14.5	12.9	29.1	41.3
GN (3)	20	1.3	4.9	4.4	7.3	12.4
	50	0.4	8.1	7.0	15.6	24.9
	100	0.8	15.0	13.7	31.2	44.1
	250	22.8	40.9	41.7	70.6	81.2
GN (5)	20	0.5	8.2	7.8	15.9	25.4
	50	0.1	23.0	24.3	47.2	61.2
	100	12.3	53.4	60.8	83.3	90.8
	250	97.0	96.7	99.1	99.9	100.0
GN (10)	20	0.3	13.1	13.8	26.4	38.5
	50	0.3	43.4	53.4	73.0	83.3
	100	50.0	84.9	94.6	97.7	99.1
	250	100.0	100.0	100.0	100.0	100.0

The remaining notation is aligned with Table 1 Best performance is marked in bold

where $[x] := \max\{k \in \mathbb{Z} : k \leq x\}$ denotes the floor of $x \in \mathbb{R}$ and

$$\bar{X}_A := \frac{1}{[n\beta] - [n\alpha]} \sum_{i=[n\alpha]+1}^{[n\beta]} X_{(i)} \tag{8}$$

is the conditional sample mean. In particular, we set $\hat{\sigma}^2 := \hat{\sigma}_{A[0,1]}^2$. Recall that the test statistic N is given by

$$N = \frac{1}{\rho} \left(\frac{\hat{\sigma}_L^2 - \hat{\sigma}_M^2}{\hat{\sigma}^2} + \frac{\hat{\sigma}_R^2 - \hat{\sigma}_M^2}{\hat{\sigma}^2} \right) \sqrt{n}, \tag{9}$$

where the normalising constant ρ in (9) is approximately equal to 1.7885; we refer to Appendix A for the closed form formula for ρ . Now, we are ready to state the main result of this section, i.e. Theorem 1.

Theorem 1 *Let $X \sim N(\mu, \sigma)$. Then,*

$$N \xrightarrow{d} \mathcal{N}(0, 1), \quad n \rightarrow \infty,$$

where N is given in (9), and ρ is a fixed normalising constant independent of μ, σ , and n .

Before we present the proof of Theorem 1 let us introduce a series of Lemmas and additional notation; proof techniques are partially based on those introduced in Stigler (1973). To ease the notation, for a fixed set A , where $A = A[\alpha, \beta]$, we define

$$\begin{aligned} \mu_A &:= \mathbb{E}[X|X \in A], & a &:= F_X^{-1}(\alpha) = \mu + \sigma \Phi^{-1}(\alpha), \\ \sigma_A^2 &:= \mathbb{E}[(X - \mu_A)^2|X \in A], & b &:= F_X^{-1}(\beta) = \mu + \sigma \Phi^{-1}(\beta), \\ \kappa_A &:= \frac{1}{(\sigma_A^2)^2} \mathbb{E}[(X - \mu_A)^4|X \in A], & m_n &:= [n\beta] - [n\alpha]. \end{aligned}$$

Additionally, we set

$$A_n := \#\{i : X_i \leq a\} = \sum_{i=1}^n \mathbb{1}_{\{X_i \leq a\}}, \quad B_n := \#\{i : X_i \leq b\} = \sum_{i=1}^n \mathbb{1}_{\{X_i \leq b\}},$$

where $\mathbb{1}_C$ is the indicator function of set C . It is useful to note that A_n and B_n follow the binomial distributions $B(n, \alpha)$ and $B(n, \beta)$, respectively; note that for $\alpha = 0$ and $\beta = 1$ the distributions are degenerate with $A_n \equiv 0$ and $B_n \equiv n$.

Finally, for any sequence (a_i) we introduce the notation of the *directed sum* that is given by

$$\mathcal{E}_{i=k}^l a_i := \begin{cases} \sum_{i=k+1}^l a_i, & \text{if } k < l, \\ 0, & \text{if } k=l, \\ -\sum_{i=l+1}^k a_i, & \text{if } k > l. \end{cases}$$

In Lemma 1, we show the consistency of the conditional sample expectation. Note that the statement of Lemma 1 does not explicitly rely on normality assumption. In fact, the proof is true under very weak conditions imposed on X (e.g. continuity of the distribution function of X); similar statement is true for other lemmas presented in this section. Also, it should be noted that Lemma 1 and 2 show consistency and asymptotic distribution of the standard non-parametric Expected Shortfall estimator; see e.g. McNeil et al. (2010) for details.

Lemma 1 For any $A = A[\alpha, \beta]$, it follows that $\bar{X}_A \xrightarrow{\mathbb{P}} \mu_A, \quad n \rightarrow \infty$.

Proof Let $A = A[\alpha, \beta]$. For any $n \in \mathbb{N}$, we get

$$\bar{X}_A = \frac{1}{m_n} \sum_{i=A_n+1}^{B_n} X(i) + \frac{1}{m_n} \mathcal{E}_{i=[n\alpha]}^{A_n} X(i) + \frac{1}{m_n} \mathcal{E}_{i=B_n}^{[n\beta]} X(i).$$

Now, we show that

$$\frac{1}{m_n} \mathcal{E}_{i=[n\alpha]}^{A_n} X(i) \xrightarrow{\mathbb{P}} 0. \tag{10}$$

Due to the consistency of the empirical quantiles, we get $X_{([n\alpha])} \xrightarrow{\mathbb{P}} a$ and $X_{(A_n)} \xrightarrow{\mathbb{P}} a$, as $n \rightarrow \infty$. Thus, using inequality

$$0 \leq \left| \frac{1}{m_n} \mathcal{E}_{i=[n\alpha]}^{A_n} X(i) \right| \leq \left| \frac{A_n - [n\alpha]}{m_n} \right| \max\{ |X_{([n\alpha])}|, |X_{(A_n)}| \},$$

to prove (10), it is sufficient to show that $\left| \frac{A_n - [n\alpha]}{m_n} \right| \xrightarrow{\mathbb{P}} 0$. Noting that

$$\frac{A_n - [n\alpha]}{m_n} = \frac{n}{m_n} \left(\frac{1}{n} A_n - \alpha \right) + \frac{n\alpha - [n\alpha]}{m_n},$$

where

$$\lim_{n \rightarrow \infty} \frac{n\alpha - [n\alpha]}{m_n} = 0, \quad \lim_{n \rightarrow \infty} \frac{n}{m_n} = \frac{1}{\beta - \alpha}, \tag{11}$$

and, by the Law of Large Numbers, $\left(\frac{1}{n} A_n - \alpha \right) \xrightarrow{\mathbb{P}} 0$, we conclude the proof of (10). The proof of

$$\frac{1}{m_n} \mathcal{E}_{i=B_n}^{[n\beta]} X(i) \xrightarrow{\mathbb{P}} 0 \tag{12}$$

is similar to the proof of (10) and is omitted for brevity.

Next, observe that

$$\frac{1}{m_n} \sum_{i=A_n+1}^{B_n} X(i) = \frac{n}{m_n} \left(\frac{1}{n} \sum_{i=1}^n X_i \mathbb{1}_{\{X_i \in A\}} \right).$$

Consequently, noting that $\mu_A = \frac{\mathbb{E}[X \mathbb{1}_{\{X \in A\}}]}{\beta - \alpha}$, and using the Law of Large Numbers, we get

$$\frac{1}{m_n} \sum_{i=A_n+1}^{B_n} X_{(i)} \xrightarrow{\mathbb{P}} \mu_A. \tag{13}$$

Combining (10), (12), and (13), we conclude the proof. □

Next, we focus on the asymptotic distribution of the conditional sample mean; note that Lemma 2 is a slight modification of the result of Stigler (1973) for trimmed means. For completeness, we present the full proof.

Lemma 2 *For any $A = A[\alpha, \beta]$, it follows that*

$$\sqrt{n} (\bar{X}_A - \mu_A) \xrightarrow{d} \mathcal{N}(0, \eta_A), \quad n \rightarrow \infty,$$

for some constant $0 < \eta_A < \infty$.

Proof For brevity, we assume that $\alpha > 0$ and $\beta < 1$. The remaining degenerate cases could be treated in the similar manner. Let $A = A[\alpha, \beta]$. Define

$$S_n := \sqrt{n} (\bar{X}_A - \mu_A).$$

As in the proof of Lemma 1, observe that

$$\begin{aligned} S_n &= \frac{\sqrt{n}}{m_n} \left(\sum_{i=A_n+1}^{B_n} X_{(i)} - m_n \mu_A + \mathcal{E}_{i=[n\alpha]}^{A_n} X_{(i)} + \mathcal{E}_{i=B_n}^{[n\beta]} X_{(i)} \right) \\ &= \frac{\sqrt{n}}{m_n} \left(\sum_{i=A_n+1}^{B_n} (X_{(i)} - \mu_A) + (A_n - [n\alpha])(a - \mu_A) + ([n\beta] - B_n)(b - \mu_A) \right. \\ &\quad \left. + \mathcal{E}_{i=[n\alpha]}^{A_n} (X_{(i)} - a) + \mathcal{E}_{i=B_n}^{[n\beta]} (X_{(i)} - b) \right). \end{aligned} \tag{14}$$

Now, we show that

$$\frac{\sqrt{n}}{m_n} \mathcal{E}_{i=[n\alpha]}^{A_n} (X_{(i)} - a) \xrightarrow{\mathbb{P}} 0. \tag{15}$$

Due to the consistency of the empirical quantiles, we get

$$(X_{([n\alpha])} - a) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad (X_{(A_n)} - a) \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$. Thus, using inequality

$$0 \leq \left| \frac{\sqrt{n}}{m_n} \mathcal{E}_{i=[n\alpha]}^{A_n} (X_{(i)} - a) \right| \leq \left| \frac{A_n - [n\alpha]}{m_n / \sqrt{n}} \right| \max\{|(X_{([n\alpha])} - a)|, |X_{(A_n)} - a|\},$$

it is sufficient to show that $\frac{A_n - [n\alpha]}{m_n/\sqrt{n}}$ converges in distribution to some non-degenerate distribution. Note that

$$\frac{A_n - [n\alpha]}{m_n/\sqrt{n}} = \frac{\sqrt{n} \left(\frac{1}{n}A_n - \alpha\right)}{m_n/n} + \frac{n\alpha - [n\alpha]}{m_n/\sqrt{n}},$$

where

$$\lim_{n \rightarrow \infty} \frac{m_n}{n} = \beta - \alpha, \quad \lim_{n \rightarrow \infty} \frac{n\alpha - [n\alpha]}{m_n/\sqrt{n}} = 0, \tag{16}$$

and, by the Central Limit Theorem applied to $A_n \sim B(n, \alpha)$, we get

$$\sqrt{n} \left(\frac{1}{n}A_n - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \sqrt{\alpha(1 - \alpha)}\right).$$

Thus, using the Slutsky’s Theorem (see e.g. (Ferguson 1996, Theorem 6’)), we get

$$\frac{A_n - [n\alpha]}{m_n/\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \sqrt{\frac{\alpha(1 - \alpha)}{\beta - \alpha}}\right),$$

which concludes the proof of (15). Similarly, one can show that

$$\frac{\sqrt{n}}{m_n} \mathcal{E}_{i=B_n}^{[n\beta]} (X_{(i)} - b) \xrightarrow{\mathbb{P}} 0. \tag{17}$$

Combining (15) with (17), and noting that

$$\frac{n\alpha - [n\alpha]}{m_n/\sqrt{n}} \rightarrow 0 \quad \text{and} \quad \frac{[n\beta] - n\beta}{m_n/\sqrt{n}} \rightarrow 0, \tag{18}$$

we can rewrite (14) as

$$S_n = \frac{\sqrt{n}}{m_n} \left(\sum_{i=A_n+1}^{B_n} (X_{(i)} - \mu_A) + (A_n - n\alpha)(a - \mu_A) + v(n\beta - B_n)(b - \mu_A) \right) + r_n,$$

where $r_n \xrightarrow{\mathbb{P}} 0$. Next, we have

$$S_n = \frac{n(\beta - \alpha)}{m_n} \left(\frac{\sqrt{n}}{n(\beta - \alpha)} \sum_{i=1}^n Z_i^A \right) + r_n,$$

where for $i = 1, \dots, n$, we set

$$Z_i^A := (X_i - \mu_A) \mathbb{1}_{\{X_i \in A\}} + (\mathbb{1}_{\{X_i \leq a\}} - \alpha)(a - \mu_A) + (\beta - \mathbb{1}_{\{X_i \leq b\}})(b - \mu_A).$$

Finally, noting that for $n \rightarrow \infty$ we get $\frac{n(\beta-\alpha)}{m_n} \xrightarrow{\mathbb{P}} 1$, and combining the Central Limit Theorem applied to (Z_i^A) with the Slutsky's Theorem we conclude the proof; note that (Z_i^A) are i.i.d. with zero mean and finite variance. \square

Next, we show that for the conditional variance estimator one can substitute the sample mean with the true mean without impacting the asymptotics. For any A , where $A = A[\alpha, \beta]$, the conditional variance estimator with known mean is given by

$$\hat{s}_A^2 := \frac{1}{m_n} \sum_{i=[n\alpha]+1}^{[n\beta]} (X_{(i)} - \mu_A)^2.$$

Lemma 3 *For any $A = A[\alpha, \beta]$, it follows that $\sqrt{n} (\hat{\sigma}_A^2 - \hat{s}_A^2) \xrightarrow{\mathbb{P}} 0, n \rightarrow \infty$.*

Proof As in the proof of Lemma 2, we focus on the case $0 < \alpha < \beta < 1$. Let $A = A[\alpha, \beta]$ and note that

$$\begin{aligned} \hat{s}_A^2 &= \frac{1}{m_n} \sum_{i=[n\alpha]+1}^{[n\beta]} (X_{(i)} - \bar{X}_A + \bar{X}_A - \mu_A)^2 \\ &= \frac{1}{m_n} \sum_{i=[n\alpha]+1}^{[n\beta]} (X_{(i)} - \bar{X}_A)^2 + (\bar{X}_A - \mu_A)^2 \\ &\quad + \frac{2}{m_n} (\bar{X}_A - \mu_A) \sum_{i=[n\alpha]+1}^{[n\beta]} (X_{(i)} - \bar{X}_A), \end{aligned}$$

where the last summand equals 0 since

$$\sum_{i=[n\alpha]+1}^{[n\beta]} X_{(i)} = m_n \bar{X}_A = \sum_{i=[n\alpha]+1}^{[n\beta]} \bar{X}_A.$$

Consequently, we get $\sqrt{n} (\hat{s}_A^2 - \hat{\sigma}_A^2) = \sqrt{n} (\bar{X}_A - \mu_A)^2$. Thus, using Lemma 1 combined with Lemma 2, we conclude the proof. \square

Now, we study the asymptotic behaviour of the conditional variance estimator; this is a key lemma that will be used in the proof of Theorem 1. Moreover, this result may be of independent interest since it allows one to construct the asymptotic confidence interval for the conditional variance.

Lemma 4 *For any $A = A[\alpha, \beta]$ it follows that*

$$\sqrt{n} (\hat{\sigma}_A^2 - \sigma_A^2) \xrightarrow{d} \mathcal{N}(0, \tau_A),$$

where

$$\begin{aligned} \tau_A^2 := & \frac{1}{(\beta - \alpha)^2} \left((\beta - \alpha)(\sigma_A^2)^2(\kappa_A - 1) \right. \\ & + \alpha(1 - \alpha) \left((a - \mu_A)^2 - \sigma_A^2 \right)^2 + \beta(1 - \beta) \left((b - \mu_A)^2 - \sigma_A^2 \right)^2 \\ & \left. - 2\alpha(1 - \beta) \left((a - \mu_A)^2 - \sigma_A^2 \right) \left((b - \mu_A)^2 - \sigma_A^2 \right) \right). \end{aligned} \tag{19}$$

Proof Due to Lemma 3, it is enough to consider \hat{s}_A^2 instead of $\hat{\sigma}_A^2$. For

$$S_n^A := \sqrt{n} \left(\hat{s}_A^2 - \sigma_A^2 \right),$$

we get

$$\begin{aligned} S_n^A = & \frac{\sqrt{n}}{m_n} \left(\sum_{i=A_n+1}^{B_n} \left((X_{(i)} - \mu_A)^2 - \sigma_A^2 \right) \right. \\ & + (A_n - [n\alpha]) \left((a - \mu_A)^2 - \sigma_A^2 \right) + ([n\beta] - B_n) \left((b - \mu_A)^2 - \sigma_A^2 \right) \\ & + \mathcal{E}_{i=[n\alpha]}^{A_n} \left((X_{(i)} - \mu_A)^2 - (a - \mu_A)^2 \right) \\ & \left. + \mathcal{E}_{i=B_n}^{[n\beta]} \left((X_{(i)} - \mu_A)^2 - (b - \mu_A)^2 \right) \right). \end{aligned} \tag{20}$$

By the arguments similar to the ones presented in the proof of Lemma 2, we get

$$\frac{\sqrt{n}}{m_n} \mathcal{E}_{i=[n\alpha]}^{A_n} \left((X_{(i)} - \mu_A)^2 - (a - \mu_A)^2 \right) \xrightarrow{\mathbb{P}} 0,$$

and

$$\frac{\sqrt{n}}{m_n} \mathcal{E}_{i=B_n}^{[n\beta]} \left((X_{(i)} - \mu_A)^2 - (b - \mu_A)^2 \right) \xrightarrow{\mathbb{P}} 0.$$

Thus, recalling (18), we can rewrite (20) as

$$\begin{aligned} S_n^A = & \frac{\sqrt{n}}{m_n} \left(\sum_{i=A_n+1}^{B_n} \left((X_{(i)} - \mu_A)^2 - \sigma_A^2 \right) \right. \\ & \left. + (A_n - n\alpha) \left((a - \mu_A)^2 - \sigma_A^2 \right) + (n\beta - B_n) \left((b - \mu_A)^2 - \sigma_A^2 \right) \right) + r_n, \end{aligned}$$

⁴Note that for degenerate cases $\alpha = 0$ and $\beta = 1$, we get $a = -\infty$ and $b = \infty$, respectively. In those cases, the convention $0 \cdot \infty = 0$ should be used.

where $r_n \xrightarrow{\mathbb{P}} 0$. Next, for $i = 1, \dots, n$, we set

$$\begin{aligned}
 Y_i^A &:= \left((X_i - \mu_A)^2 - \sigma_A^2 \right) \mathbb{1}_{\{X_i \in A\}} \\
 &\quad + \left(\mathbb{1}_{\{X_i \leq a\}} - \alpha \right) \left((a - \mu_A)^2 - \sigma_A^2 \right) \\
 &\quad + \left(\beta - \mathbb{1}_{\{X_i \leq b\}} \right) \left((b - \mu_A)^2 - \sigma_A^2 \right),
 \end{aligned}
 \tag{21}$$

and by straightforward computations get $\mathbb{E}[Y_i^A] = 0$ and $D^2[Y_i^A] = (\beta - \alpha)^2 \tau_A^2$. Consequently, noting that

$$S_n^A = \frac{\sqrt{n}}{m_n} \sum_{i=1}^n Y_i^A + r_n,
 \tag{22}$$

and using the Central Limit Theorem combined with the Slutsky’s Theorem, we conclude the proof. □

Finally, we are ready to show the proof of Theorem 1.

Proof (of Theorem 1) For conditioning sets L, M , and R given in (5), we define the associated sequences of random variables $(Y_i^L), (Y_i^M), (Y_i^R)$ using (21). For any $n \in \mathbb{N}$, we set

$$Z_n := \sqrt{n} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n \frac{1}{\tilde{q}} Y_i^L \\ \frac{1}{n} \sum_{i=1}^n \frac{1}{1-2\tilde{q}} Y_i^M \\ \frac{1}{n} \sum_{i=1}^n \frac{1}{\tilde{q}} Y_i^R \end{bmatrix},$$

where \tilde{q} is defined via (6). By the multivariate Central Limit Theorem (cf. (Ferguson 1996, Theorem 5)), we get $Z_n \xrightarrow{d} \mathcal{N}_3(0, \Sigma)$, where

$$\Sigma := \begin{bmatrix} \frac{\text{Cov}(Y_1^L, Y_1^L)}{\tilde{q}^2} & \frac{\text{Cov}(Y_1^M, Y_1^L)}{\tilde{q}(1-2\tilde{q})} & \frac{\text{Cov}(Y_1^R, Y_1^L)}{\tilde{q}^2} \\ \frac{\text{Cov}(Y_1^L, Y_1^M)}{\tilde{q}(1-2\tilde{q})} & \frac{\text{Cov}(Y_1^M, Y_1^M)}{(1-2\tilde{q})^2} & \frac{\text{Cov}(Y_1^R, Y_1^M)}{\tilde{q}(1-2\tilde{q})} \\ \frac{\text{Cov}(Y_1^L, Y_1^R)}{\tilde{q}^2} & \frac{\text{Cov}(Y_1^M, Y_1^R)}{\tilde{q}(1-2\tilde{q})} & \frac{\text{Cov}(Y_1^R, Y_1^R)}{\tilde{q}^2} \end{bmatrix}.$$

Now, let $S_n := \sqrt{n} (\hat{\sigma}_L^2 + \hat{\sigma}_R^2 - 2\hat{\sigma}_M^2)$. Using (2), it is easy to see that

$$S_n = \sqrt{n} \left(\hat{\sigma}_L^2 - \sigma_L^2 + \hat{\sigma}_R^2 - \sigma_R^2 - 2\hat{\sigma}_M^2 + 2\sigma_M^2 \right).
 \tag{23}$$

Consequently, by the arguments similar to the ones presented in the proof of Lemma 4 (see (22)), we can rewrite (23) as $S_n = M_n Z_n + r_n$, where $r_n \xrightarrow{\mathbb{P}} 0$ and

$$M_n := \left[\frac{n\tilde{q}}{[n\tilde{q}]}, -2 \frac{n(1-2\tilde{q})}{[n(1-\tilde{q})]-[n\tilde{q}]}, \frac{n\tilde{q}}{n-[n(1-\tilde{q})]} \right].$$

Next, observing that $M_n \xrightarrow{\mathbb{P}} [1, -2, 1]$ and using the multivariate Slutsky’s Theorem (cf. (Ferguson 1996, Theorem 6)), we get $S_n \xrightarrow{d} \mathcal{N}(0, \tau)$, where

$$\tau := \sqrt{[1, -2, 1] \Sigma [1, -2, 1]^T}. \tag{24}$$

Let $\rho := \frac{\tau}{\sigma^2}$ and $N_n := \frac{1}{\rho} \frac{S_n}{\hat{\sigma}_n^2}$. Observing that $\sigma^2/\hat{\sigma}_n^2 \xrightarrow{\mathbb{P}} 1$, and again using the Slutsky’s Theorem, we get $N_n \xrightarrow{d} \mathcal{N}(0, 1)$. To conclude the proof of Theorem 1, we need to show that ρ is independent of μ and σ .

To do so, let us first show that for any A , where $A = A[\alpha, \beta]$, and the corresponding random variable Y_1^A given in (21), we get

$$Y_1^A = \sigma^2 \psi(\tilde{X}_1, \alpha, \beta), \tag{25}$$

where $\tilde{X}_1 := (X_1 - \mu)/\sigma$ and $\psi: \mathbb{R} \times [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ is some fixed measurable function. From (Johnson et al. 1994, Sect. 13.10.1), we know that

$$\mu_A = \sigma \frac{\phi(\Phi^{-1}(\alpha)) - \phi(\Phi^{-1}(\beta))}{\beta - \alpha} + \mu, \tag{26}$$

$$\sigma_A^2 = \sigma^2 \left(\frac{\Phi^{-1}(\alpha)\phi(\Phi^{-1}(\alpha)) - \Phi^{-1}(\beta)\phi(\Phi^{-1}(\beta))}{\beta - \alpha} - \frac{(\phi(\Phi^{-1}(\alpha)) - \phi(\Phi^{-1}(\beta)))^2}{(\beta - \alpha)^2} + 1 \right).^5 \tag{27}$$

Consequently, the standardised mean $\tilde{\mu}_A := \frac{\mu_A - \mu}{\sigma}$ and variance $\tilde{\sigma}_A^2 := \frac{\sigma_A^2}{\sigma^2}$ depend only on α and β . Recalling (21), we get

$$\begin{aligned} \frac{Y_1^A}{\sigma^2} &= \left(\left(\frac{X_1 - \mu_A}{\sigma} \right)^2 - \tilde{\sigma}_A^2 \right) \mathbb{1}_{\{X_1 \in A\}} + (\mathbb{1}_{\{X_1 \leq a\}} - \alpha) \left(\left(\frac{a - \mu_A}{\sigma} \right)^2 - \tilde{\sigma}_A^2 \right) \\ &\quad + (\beta - \mathbb{1}_{\{X_1 \leq b\}}) \left(\left(\frac{b - \mu_A}{\sigma} \right)^2 - \tilde{\sigma}_A^2 \right) \\ &= \left((\tilde{X}_1 - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right) \mathbb{1}_{\{X_1 \in A\}} + (\mathbb{1}_{\{X_1 \leq a\}} - \alpha) \left((\Phi^{-1}(\alpha) - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right) \\ &\quad + (\beta - \mathbb{1}_{\{X_1 \leq b\}}) \left((\Phi^{-1}(\beta) - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right). \tag{28} \end{aligned}$$

⁵For $\alpha = 0$ or $\beta = 1$ the convention $0 \cdot \pm\infty = 0$ is used.

Combining this with equalities

$$\begin{aligned} \{X_1 \in A\} &= \{\tilde{X}_1 \in [\Phi^{-1}(\alpha), \Phi^{-1}(\beta)]\}, \\ \{X_1 \leq a\} &= \{\tilde{X}_1 \leq \Phi^{-1}(\alpha)\}, \\ \{X_1 \leq b\} &= \{\tilde{X}_1 \leq \Phi^{-1}(\beta)\}, \end{aligned}$$

we conclude the proof of (25).

Now, using (25) for $L, M,$ and $R,$ and expressing Σ/σ^4 as

$$\begin{bmatrix} \frac{\text{Cov}\left(\frac{y_1^L}{\sigma^2}, \frac{y_1^L}{\sigma^2}\right)}{\tilde{q}^2} & \frac{\text{Cov}\left(\frac{y_1^M}{\sigma^2}, \frac{y_1^L}{\sigma^2}\right)}{\tilde{q}(1-2\tilde{q})} & \frac{\text{Cov}\left(\frac{y_1^R}{\sigma^2}, \frac{y_1^L}{\sigma^2}\right)}{\tilde{q}^2} \\ \frac{\text{Cov}\left(\frac{y_1^L}{\sigma^2}, \frac{y_1^M}{\sigma^2}\right)}{\tilde{q}(1-2\tilde{q})} & \frac{\text{Cov}\left(\frac{y_1^M}{\sigma^2}, \frac{y_1^M}{\sigma^2}\right)}{(1-2\tilde{q})^2} & \frac{\text{Cov}\left(\frac{y_1^R}{\sigma^2}, \frac{y_1^M}{\sigma^2}\right)}{\tilde{q}(1-2\tilde{q})} \\ \frac{\text{Cov}\left(\frac{y_1^L}{\sigma^2}, \frac{y_1^R}{\sigma^2}\right)}{\tilde{q}^2} & \frac{\text{Cov}\left(\frac{y_1^M}{\sigma^2}, \frac{y_1^R}{\sigma^2}\right)}{\tilde{q}(1-2\tilde{q})} & \frac{\text{Cov}\left(\frac{y_1^R}{\sigma^2}, \frac{y_1^R}{\sigma^2}\right)}{\tilde{q}^2} \end{bmatrix},$$

we see that Σ/σ^4 does not depend on μ and $\sigma.$

Finally, recalling (24) and the definition of ρ we conclude the proof of Theorem (1); we refer to Appendix A for the closed-form formula for $\rho.$ □

The results presented in this section could be directly applied to various other non-parametric quantile estimators and to the unbiased variance estimators.⁶ This is summarised in the next two remarks.

Remark 1 The standard formula for the whole sample (unbiased) variance uses $n - 1$ instead of n in the denominator. In the conditional case, this would be reflected in the different formula for (7), where m_n is replaced by $m_n - 1.$ Note that the statement of Theorem 1 remains valid for the modified conditional variance estimator due to the combination of the Slutsky’s Theorem and the fact that $(m_n - 1)/m_n \rightarrow 1.$

Remark 2 When defining the conditional sample variance (7), we used $[n\alpha] + 1$ and $[n\beta]$ as the limits of the summation in (7) and (8). This choice corresponds to the non-parametric α -quantile estimator given by $X_{([n\alpha])}.$

In the literature there exist many different formulas for non-parametric quantile estimators, most of which are bounded by the nearest order statistics; see Hyndman and Fan (1996) for details. It is relatively easy to show that all results presented in this section hold true if we replace $[n\alpha]$ and $[n\beta]$ by suitably chosen sequences that correspond to different empirical quantile choices. For completeness, we provide a more detailed description of this statement.

Consider sequences (α_n) and (β_n) such that $n\alpha - \alpha_n$ and $\beta_n - n\beta$ are bounded, and define $\tilde{m}_n := \beta_n - \alpha_n.$ The corresponding conditional sample mean and variance are given by

⁶ In particular, this refers to the estimator implemented via *quantile* function in **R** that was used in Fig. 1.

$$\bar{X}_A^* := \frac{1}{\tilde{m}_n} \sum_{i=\alpha_n+1}^{\beta_n} X_{(i)} \quad \text{and} \quad \hat{\sigma}_A^{2,*} := \frac{1}{\tilde{m}_n} \sum_{i=\alpha_n+1}^{\beta_n} (X_{(i)} - \bar{X}_A^*)^2.$$

Then, we can replace \bar{X}_A and $\hat{\sigma}_A^2$ by \bar{X}_A^* and $\hat{\sigma}_A^{2,*}$ in Theorem 1 as well as in all lemmas presented in the section.

Instead of showing a full proof, we briefly comment how to show consistency of quantile estimators as well as comment on counterparts of (11) and (16). All proofs could be translated using a very similar logic.

First, noting that for some $k \in \mathbb{N}$ we get $X_{([n\alpha]-k)} \leq X_{(\alpha_n)} \leq X_{([n\alpha]+k)}$ and $X_{([n\beta]-k)} \leq X_{(\beta_n)} \leq X_{([n\beta]+k)}$, it is straightforward to check that $X_{(\alpha_n)}$ and $X_{(\beta_n)}$ are consistent α -quantile and β -quantile estimators; see e.g. (Serfling 1980, Sect. 2.3). Second, to show the analogue of (11), it is enough to use the boundedness of $n\alpha - \alpha_n$ and $\beta_n - n\beta$, and note that $\frac{n\alpha - \alpha_n}{n} \rightarrow 0$ and $\frac{\beta_n - n\beta}{n} \rightarrow 0$. Third, to show (16), it is enough to use boundedness of $n\alpha - \alpha_n$ and note that for some $k \in \mathbb{N}$ we get $\frac{|n\alpha - \alpha_n|}{\tilde{m}_n/\sqrt{n}} \leq \frac{k}{\tilde{m}_n} \sqrt{n} = \frac{\tilde{m}_n}{n} \frac{k}{\sqrt{n}}$.

6 Empirical example: case study of market stock returns

In this section we apply the proposed framework to stock market returns performing a basic sanity-check verification. Before we do that, let us comment on the connection between the 20-60-20 Rule and financial time series.

Assuming that X describes financial asset return rates we can split the population using 20/60/20 ratio and check the behaviour of returns within each subset. If non-normal perturbations are observed only for extreme events, the 20/60/20 break might identify the regime switch and provide a good spatial clustering; this could be linked to a popular financial *stylised fact* saying that average financial asset returns tend to be normal, but the extreme returns are not—see Cont (2001) and Sheikh and Qiao (2010) for details. It should be emphasized that according to authors’ best knowledge, the link between this property and data non-normality was not discussed in the literature before.

The easiest way to verify this hypothesis is to take stock return samples for different periods, make the quantile–quantile plots (with standard normal as a reference distribution) and check if the clustering is accurate. In Fig. 4, we present exemplary results for two major US stocks, namely GOOGL and AAPL, and two major stock indices, namely S&P500 and DAX; we took time-series of length 250 for different time intervals ranging the period from 10/2015 to 01/2018.⁷

From Fig. 4 we see that this division is surprisingly accurate: a very good normal fit is observed in the M set (middle 60% of observations), while the fit in the tail sets L and R (bottom and top 20% of observations) is bad. By taking different sample sizes, different time-horizons, and different stocks we can confirm that this property is systematic, i.e. the results are almost always similar to the ones presented in Fig. 4.

⁷ Data is downloaded from [Yahoo Finance](#) via **R** *tidyquant* package.

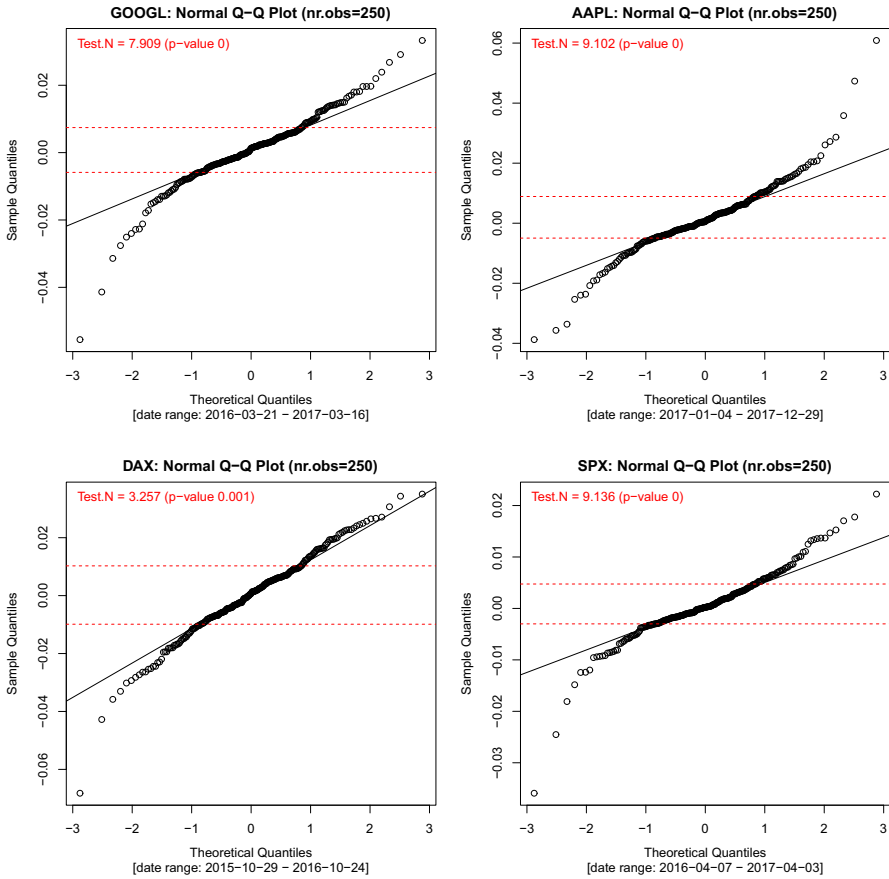


Fig. 4 Quantile–Quantile plots for various financial returns of size $n = 250$. The dashed lines correspond to 20% and 80% quantiles. One can see different data behaviour in each cluster: while the central region is aligned with normal distribution the tail regions are not

While the presence of fat-tails in asset return distributions is a well-known observation in the financial world, it is quite surprising to note that the non-normal behaviour could be seen for approximately 40% of data. Also, test statistic N can be used to formally quantify this phenomenon and to measure tail heaviness: the bigger the conditional standard deviation in the tails (in reference to the central part), the fatter the tails.

In the following, we focus on assessing the performance of the test statistic N on market data. We perform a simple empirical study and take returns of all stocks listed in S&P500 index on 16.06.2018 that have full historical data in the period from 01.2000 to 05.2018. This way we get full data (4610 daily adjusted close price returns) for 381 stocks. Next, for a given sample size $n \in \{50, 100, 250\}$ we split the returns into disjoint sets of length n , and for each subset we compare the value of N with the corresponding empirical quantiles presented in Fig. 2. More precisely, using N we perform a right-

Table 4 The table contains results of performance tests on the market data

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

Column *Desc* gives the name of performance measure, *nr runs* indicates how many subsamples were created for a given *n*, the value α gives the confidence level, and column *rejects* gives ratio of subsample for which at least one test rejected the normality assumption. The acronyms JB, AD, SW, and N refer to Jarque–Bera Test, Anderson–Darling Test, Shapiro–Wilk Test, and *N* test, respectively
Best performance is marked in bold

sided statistical test and reject normality (null) hypothesis if the computed value is greater than the empirical value $F_n^{-1}(1 - \alpha)$, for $\alpha \in \{1\%, 2.5\%, 5\%\}$.

To assess test performance, we compare the results with other benchmark normality tests: Jarque–Bera test, Anderson–Darling test, and Shapiro–Wilk test. While the non-normality of returns is a well-known fact, and all testing frameworks should show good performance, we want to check if our framework leads to some new interesting results. We check the normality hypothesis and compute two supplementary metrics that are used for performance assessment:

- **Statistic T** gives the **total rejection ratio** of a given test. It corresponds to the proportion of data on which the normality assumption was rejected at a given significance level; it is the ratio of rejected data subsamples to all data subsamples.
- **Statistic U** gives the **unique rejection ratio** of a given test. It corresponds to the proportion of data on which normality assumption was rejected at a given significance level only by the considered test (among all four tests); it is the ratio of uniquely rejected data subsamples to all data subsamples.

The combined results for all values of *n* and α are presented in Table 4.

One can see that the statistic *N* performs very well and gives the best results for all choices of *n*. Surprisingly, our testing framework allows one to detect non-normal

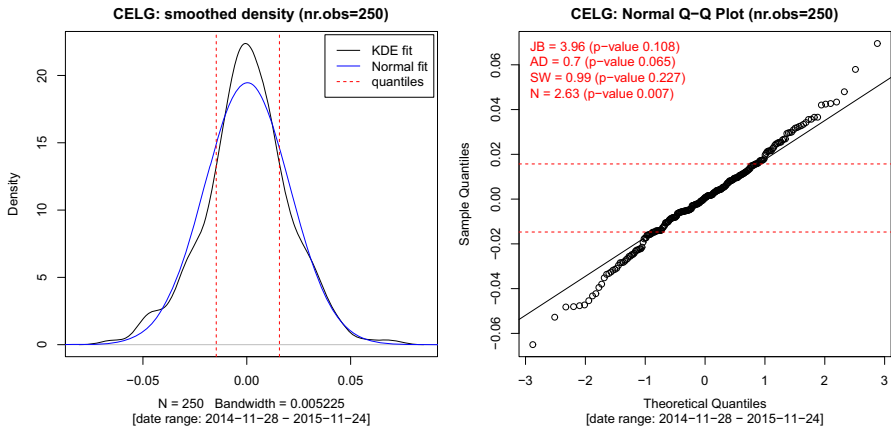


Fig. 5 Exemplary time series for which only test N rejected normality at level 1%. KDE fit corresponds to empirical density obtained with Kernel Density Estimation, while normal fit correspond to a standard normal fit. Empirical quantiles corresponding to 20/60/20 ratio are marked with dashed line

behaviour in cases when other tests fail: the outcomes of measure U are material in all cases. For example, for $n = 50$ and $\alpha = 5\%$, the value of U was equal to 5.1%—this corresponds to almost 11% of all rejected samples. The results are especially striking for $n = 250$, where the normality assumption was rejected in almost all cases (ca. 90%). While one might think that for such a big sample size the three classical tests should detect all abnormalities, our test still uniquely rejected normality in multiple cases. For $\alpha = 1\%$, the normality was rejected for additional 262 samples (3.8% of the population). For transparency, in Fig. 5 we show exemplary data subset for which this happened.

7 Concluding remarks and other applications

In this paper we have shown that the test statistic N introduced in (3) could be used to measure the heaviness of the tails in reference to the central part of distribution and could serve as an efficient goodness-of-fit normality test statistic. Test statistic N is based on the conditional second moments, performs quite well on market financial data, and allows one to detect non-normal behaviour where other benchmark tests fail.

As mentioned in the introduction, most empirical studies suggest that the normality tests should be chosen carefully as their statistical power varies depending on the context. Our proposal proves to have the best test power in the cases when the true distribution is assumed to be symmetric and have tails that are fatter or slimmer than the normal one. It should be noted that our test is in fact based on the implicit distribution symmetry assumption. Indeed, in (3), the impact of left and right tail is taken with the same weight. Nevertheless, this could be easily generalised e.g. by considering only one of the tail variances; we comment on that later.

In Theorem 1 we proved that the asymptotic distribution of N is normal under the normality null hypothesis. This allows us to study the shape of rejection intervals for

sufficiently large samples. To obtain this result, in Lemma 4 we derived the asymptotic distribution of the conditional sample variance.

Also, we showed that the 20-60-20 Rule explains the financial *stylised fact* related to tail non-normal behaviour and provides surprisingly accurate clustering of asset return time series. Quite surprisingly, non-normality is visible for almost 40% of the observations.

In summary, we believe that tail-impact tests based on the conditional second moments are very promising and provide a nice alternative to the classical framework based e.g. on the third and fourth moments.

For example, the multivariate extension of the test statistic N could be defined using the results presented in Jaworski and Pitera (2016), e.g. to assess the adequacy of using the correlation structure for dependence modeling. Also, this could be extended to any multivariate elliptic distribution using the results from Jaworski and Pitera (2017).

The construction of N shows how to use conditional second moments for statistical purposes. In fact, one might introduce various other statistics that test underlying distributional assumptions. Let us present a couple of examples:

- We can test only the (left) low-tail impact on the central part by considering one of test statistics

$$N_1 := \left(\frac{\hat{\sigma}_L^2 - \hat{\sigma}_M^2}{\hat{\sigma}^2} \right) \sqrt{n}, \quad N_2 := \left(\frac{\hat{\sigma}_L^2 - \hat{\sigma}_M^2}{\hat{\sigma}_M^2} \right) \sqrt{n}.$$

- For any quantile-based conditioning sets A and B , and any elliptical distribution, one can introduce the statistic

$$N_3 := \left(\frac{\hat{\sigma}_A^2}{\hat{\sigma}_B^2} - \lambda \right) \sqrt{n},$$

where $\lambda \in \mathbb{R}$ is a constant depending on the quantiles that define conditioning sets and the underlying distribution. Assuming that $A = L$ and $B = \mathbb{R}$ (whole space), we get the proportion between the tail dispersion and overall dispersion. In this specific case, in the normal framework, we get

$$\lambda = 1 - \frac{\Phi^{-1}(0.2)\phi(\Phi^{-1}(0.2))}{0.2} - \frac{(\phi(\Phi^{-1}(0.2)))^2}{0.2^2};$$

see (Jaworski and Pitera 2016, Sect. 3) for details.

Note that under the normality assumption all proposed statistics are pivotal quantities which facilitates an easy and efficient hypothesis testing; the asymptotic distribution for all statistics could be derived using similar reasoning as the one presented in Theorem 1.

Acknowledgements Jagiellonian University in Krakow.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Closed-form formula for the normalising constant

In this section, we present the closed-form formula for the normalising constant ρ from Theorem 1. For brevity, we omit detailed calculations and only present the outcome.

To ease the notation, for any $\gamma \in [0, 1]$, we set $x_\gamma := \Phi^{-1}(\gamma)$. Then, for any $A = A[\alpha, \beta]$, the standardised second, third, and fourth conditional central moments are given by⁸

$$\begin{aligned}
 m_A^{(2)} &:= 1 + \frac{x_\alpha \phi(x_\alpha) - x_\beta \phi(x_\beta)}{\beta - \alpha}, \\
 m_A^{(3)} &:= \frac{(x_\alpha)^2 \phi(x_\alpha) - (x_\beta)^2 \phi(x_\beta)}{\beta - \alpha} + 2 \frac{\phi(x_\alpha) - \phi(x_\beta)}{\beta - \alpha}, \\
 m_A^{(4)} &:= 3 + \frac{(x_\alpha)^3 \phi(x_\alpha) - (x_\beta)^3 \phi(x_\beta)}{\beta - \alpha} + 3 \frac{x_\alpha \phi(x_\alpha) - x_\beta \phi(x_\beta)}{\beta - \alpha}.
 \end{aligned}$$

Moreover, the standardised conditional mean, conditional variance, and conditional kurtosis are equal to

$$\begin{aligned}
 \tilde{\mu}_A &= \frac{\phi(x_\alpha) - \phi(x_\beta)}{\beta - \alpha}, \\
 \tilde{\sigma}_A^2 &= m_A^{(2)} - (\tilde{\mu}_A)^2, \\
 \tilde{\kappa}_A &= \frac{m_A^{(4)} - 3(\tilde{\mu}_A)^4 + 6(\tilde{\mu}_A)^2 m_A^{(2)} - 4\tilde{\mu}_A m_A^{(3)}}{(\tilde{\sigma}_A^2)^2}.
 \end{aligned}$$

Also, recall that $\tilde{q} = \Phi(x)$, where x is the unique negative solution of the equation

$$-x\Phi(x) - \phi(x)(1 - 2\Phi(x)) = 0.$$

Now, we are ready to present the closed-form formula for ρ ; see Corollary 1.

Corollary 1 *The normalising constant ρ from Theorem 1 is given by*

$$\rho := \sqrt{\frac{\tau_L^2}{\sigma^4} + 4 \frac{\tau_M^2}{\sigma^4} + \frac{\tau_R^2}{\sigma^4} - \frac{4(C_1 + C_2)}{\tilde{q}(1 - 2\tilde{q})} + \frac{2C_3}{\tilde{q}^2}},$$

⁸ Recall that for $\alpha = 0$ or $\beta = 1$ we follow the convention $0 \cdot \pm\infty = 0$.

where for $A \in \{L, M, R\}$ we have

$$\begin{aligned} \frac{\tau_A^2}{\sigma^4} &= \frac{1}{(\beta - \alpha)^2} \left((\beta - \alpha)(\tilde{\sigma}_A^2)^2(\tilde{\kappa}_A - 1) + \alpha(1 - \alpha) \left((x_\alpha - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right)^2 \right. \\ &\quad \left. + \beta(1 - \beta) \left((x_\beta - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right)^2 \right. \\ &\quad \left. - 2\alpha(1 - \beta) \left((x_\alpha - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right) \left((x_\beta - \tilde{\mu}_A)^2 - \tilde{\sigma}_A^2 \right) \right), \end{aligned}$$

and constants C_1 , C_2 , and C_3 are given by

$$\begin{aligned} C_1 &= \tilde{q}^2 \left((x_{\tilde{q}} - \tilde{\mu}_L)^2 - \tilde{\sigma}_L^2 \right) \left((x_{\tilde{q}} + \tilde{\mu}_M)^2 - \tilde{\sigma}_M^2 \right) \\ &\quad - \tilde{q}(1 - \tilde{q}) \left((x_{\tilde{q}} - \tilde{\mu}_M)^2 - \tilde{\sigma}_M^2 \right) \left((x_{\tilde{q}} - \tilde{\mu}_L)^2 - \tilde{\sigma}_L^2 \right), \\ C_2 &= \tilde{q}^2 \left((x_{\tilde{q}} + \tilde{\mu}_R)^2 - \tilde{\sigma}_R^2 \right) \left((x_{\tilde{q}} - \tilde{\mu}_M)^2 - \tilde{\sigma}_M^2 \right) \\ &\quad - \tilde{q}(1 - \tilde{q}) \left((x_{\tilde{q}} + \tilde{\mu}_M)^2 - \tilde{\sigma}_M^2 \right) \left((x_{\tilde{q}} + \tilde{\mu}_R)^2 - \tilde{\sigma}_R^2 \right), \\ C_3 &= -\tilde{q}^2 \left((x_{\tilde{q}} + \tilde{\mu}_R)^2 - \tilde{\sigma}_R^2 \right) \left((x_{\tilde{q}} - \tilde{\mu}_L)^2 - \tilde{\sigma}_L^2 \right). \end{aligned}$$

Approximately, the value of ρ is equal to 1.7885.

References

- Alexander C (2009) Market risk analysis, quantitative methods in finance, vol I. Wiley, Hoboken
- Anderson TW, Darling DA (1954) A test of goodness of fit. J Am Stat Assoc 49(268):765–769
- Brockwell P, Davis R (2016) Introduction to time series and forecasting. Springer, New York
- Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quant Finan 1(2):223–236
- Desgagné A, Lafaye de Micheaux P (2018) A powerful and interpretable alternative to the Jarque-Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family. J Appl Stat 45(13):2307–2327
- Ferguson T (1996) A course in large sample theory. Springer, Dordrecht
- Hair J, Black W, Babin B, Anderson R (2013) Multivariate data analysis: Pearson new, international edn. Pearson Education Limited, London
- Henze N (2002) Invariant tests for multivariate normality: a critical review. Stat Pap 43(4):467–506
- Hyndman RJ, Fan Y (1996) Sample quantiles in statistical packages. Am Stat 50(4):361–365
- Jarque CM, Bera AK (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econom Lett 6(3):255–259
- Jaworski P, Pitera M (2016) The 20–60–20 rule. DISCRETE CONT DYN-B 21(4)
- Jaworski P, Pitera M (2017) A note on conditional covariance matrices for elliptical distributions. Stat Probabil Lett 129:230–235
- Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1. Wiley, New York
- Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis, vol 344. Wiley, Hoboken
- Madansky A (2012) Prescriptions for working statisticians. Springer, New York
- McNeil AJ, Frey R, Embrechts P (2010) Quantitative risk management: concepts, techniques, and tools. Princeton University Press, Princeton
- Nadarajah S (2005) A generalized normal distribution. J Appl Stat 32(7):685–694

- Romão X, Delgado R, Costa A (2010) An empirical power comparison of univariate goodness-of-fit tests for normality. *J Statl Comput Simul* 80(5):545–591
- Romesburg C (2004) Cluster analysis for researchers. Lulu Press, Morrisville
- Serfling R (1980) Approximation theorems of mathematical statistics. Wiley series in probability and mathematical statistics. Wiley, New York
- Sheikh A, Qiao H (2010) Non-normality of market returns: a framework for asset allocation decision making. *J Altern Invest* 12(3):8–35
- Stigler SM (1973) The asymptotic distribution of the trimmed mean. *Ann Stat* 1(3):472–477
- Thadewald T, Büning H (2007) Jarque-Bera test and its competitors for testing normality—a power comparison. *J Appl Stat* 34(1):87–105
- Thode HC (2002) Testing for normality. Marcel Dekker, New York
- Tumlinson S, Keating J, Balakrishnan N (2016) Linear estimation for the extended exponential power distribution. *J Stat Comput Simul* 86(7):1392–1403
- Wilk MB, Shapiro SS (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52(3–4):591–611

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.