Robust Tests for the Mean for Heavy-Tailed Data*

Samuel P. Engle†

University Exeter

Department of Economics

This version: November 24, 2022

Please find the current version here

Abstract

The t-test is a standard inferential procedure in economics and finance. When the data exhibit heavy tails, the t-test may have low power. This paper makes two contributions. First, the rate at which power converges to 1 for data in a particular class of heavy-tailed distributions is characterized. While classical results on the rate of convergence of power focus on exponential rates, we find the rate to be a much slower polynomial rate when the data have heavy tails. Second, a new testing procedure is developed which improves upon the rates in the class of heavy-tailed distributions under consideration. Simulation evidence shows that the efficiency gains from the new testing procedure can be substantial.

1 Introduction

Since its introduction in Student (1908), the usual t-test for inference about the mean has played a ubiquitous role in theory and practice in econometrics and statistics. Initially motivated as the optimal test in the canonical inference problem with Gaussian observations, asymptotic arguments have lead to the application of the t-test to scenarios in which the data are not normally distributed. Heavy-tailed data are a particular departure from normality that has been of increasing interest. This paper develops new results characterizing the

*I am grateful for helpful feedback and comments I have received from Jack Porter, Bruce Hansen, Mikkel Sølvsten, Harold Chiang, Giuseppe Cavaliere, Annie, Lee, Anson Zhou, Anna Trubnikova, John Stromme, and other seminar participants at the University of Wisconsin-Madison and the University of Exeter. All remaining errors are my own.

†email: s.p.engle@exeter.ac.uk website: samuelpengle.com

power of the t-test when the underlying distributions have heavy tails. Furthermore, we develop an alternative to the t-test which provides consistent inference for the mean, and provides efficiency gains for heavy-tailed data.

The first main contribution of this paper is establishing the rate at which the power of the t-test converges to 1 under a fixed alternative when the data exhibit heavy tails. In classical settings, when the moment generating function exists the type-II error disappears at an exponential rate in the sample size for any fixed alternative. We show that when the moment generating function does not exist, under a different set of regularity conditions type-II error disappears at a much slower rate. The worst-case rate in the class we consider is a polynomial rate in sample size.

Our results complement existing results in the statistical and econometric literature on using the t-test with heavy-tailed data. Recently, Müller (2019) and Müller (2020) establish slow rates of convergence of the t-test statistic to a standard normal random variable under the null hypothesis when the data exhibit Pareto-like tails. We find that similar slow-convergence results hold when looking at the type-II error rate under a fixed alternative. Shephard (2020) proposes an alternative estimator in regression settings with heavy-tailed data to deal with similar issues. Young (2021) argues that not only are heavy-tailed data prevalent in economic applications, but heteroskedastic-robust inference is particularly sensitive to heavy-tailed data.

The second key contribution of this paper is the construction a robust test statistic which avoids the slow convergence rate of the Type-II error rate exhibited by the t-test. The new testing procedure preserves the asymptotic local power properties of the t-test while providing for a faster rate of convergence of the Type-II error rate to zero. Our strategy is based on robust estimation methods in machine learning. There is a growing literature on constructing efficient, robust estimators of the mean. This has been a departure from classical statistics, where robust methods focused on using alternative moment conditions to estimate a parameter of interest. Generally, there are two approaches. One approach, such as in Sun et al. (2020), is to use an adapted procedure based on a sequence of robust estimation problems. An alternative approach is to use sample splitting and recombine sub-sample estimators in a robust way. This has been done in Brownlees et al. (2015), Minsker (2019), and Mathieu and Minsker (2021), among others. Namely, we adapt and extend results in Minsker (2022) for robust estimation of the mean to construct a test statistic based on sample splitting and robust recombination of the data. We contribute to this literature by showing that the key condition required for asymptotic normality under these robust recombination procedures is an asymptotic quadratic mean differentiability condition.

We also contribute to a long history in the statistical literature on the properties of the t-test statistic in heavy-tailed settings. Under symmetry conditions, it is shown in Efron (1969) that the t-test will tend to be asymptotically conservative when the tails of the data are sufficiently heavy. In Giné et al. (1997), necessary and sufficient conditions for the t-test statistic to be asymptotically standard-normal or subgaussian are provided. In Shao (1999) large-deviation results for the t-test statistic are established.

This paper also contributes to the literature on asymptotic efficiency of the t-test. Hodges and Lehmann (1956) was the first paper to propose the relative efficiency measure we adopt in this paper, and they derived the efficiency of the t-test when the observations are normally distributed. Recently, He and Shao (1996) derived the Bahadur efficiency of closely-related normalized score tests. Their results show that t-tests are reasonably robust to heavy-tails when considering the behavior of p-values. Our results show that heavy-tails lead to slow convergence of the type-II error rate to zero, and therefore provide a new and different perspective on the efficiency properties of the t-test.

For this paper we focus on the cases where we have an i.i.d. sample $X_1, \ldots, X_n$, with $\mathbb{E}\, X_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. The tests we consider are hypothesis tests of the form:

$$H_0 : \mu = \mu_0 \qquad \text{v.s.} \qquad H_1 : \mu > \mu_0 \tag{1}$$

We will mainly focus on the behavior of the t-test, however it will be useful for exposition to also define a test where the variance is known and used. In each case the null hypothesis is rejected for sufficiently large values of the test statistic

$$Z_n := \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \qquad T_n := \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \tag{2}$$

where $\bar{X}$ is the sample mean and S^2 is the sample variance. T_n is the classic t-test statistic, and Z_n is what we will refer to as the z-test statistic. In both cases, we reject the null hypothesis when the test statistic is sufficiently large. Note that the z-test statistic is generally not available, however relative to the t-test statistic, the properties of Z_n are easier to derive.

This paper proceeds as follows. In Section 2, we present our main result on the efficiency of the t-test and compare with other common relative efficiency results in this setting. In Section 3 we discuss how to construct robust tests. In Section 4 we provide some simulation evidence. In Section 5 we conclude. The proofs of the main results are in Appendix A, and supporting technical results are in Appendix B.

2 Efficiency of the t-test

We first present the main result of this paper: a characterization of the asymptotic type-II error rate for tests T_n when the observations have heavy tails. We then compare these results to relative efficiency comparisons based on local asymptotic power and Bahadur relative efficiency.

2.1 Hodges-Lehmann Relative Efficiency

The Neyman-Pearson testing paradigm evaluates tests of (1) by valuing tests which have a low probability of rejecting H_0 when H_1 is true, subject to a constraint on how often H_0 is rejected when H_0 is true. Put another way, in general the aim is to minimize type-II errors while constraining the type-I error rate. In settings where the observations X_i belong to a parametric family, finite-sample optimality of tests has been thoroughly explored; see Lehmann and Romano (2005) for a survey. In many applications researchers do not want to assume the observations belong to a particular parametric family, and would like to conduct semiparametric inference on a finite-dimensional functional of the distribution of the observations. In this paper we focus on the mean as such a summary of interest. In these semiparametric settings, practitioners focus their attention on tests which are asymptotically valid. Most conventional tests result in the type-II error rate converging to 0 asymptotically, presenting an inherent challenge in comparing tests which can primarily be characterized with asymptotic methods.

When comparing two testing procedures, the natural mode is relative efficiency. Let us consider two tests $\phi_{1,n}$ and $\phi_{2,n}$ for testing (1). For a given Type-I error rate α, and alternative μ, let $h(n)$ be the smallest sample size n' such that if $\phi_{1,n} = 1 - \beta$ at μ, then $\phi_{2,n'} \geq 1 - \beta$. Comparing the ratio $h(n)/n$ when $\mu \to \mu_0$ as $n \to \infty$ leads to the relative efficiency comparison of Pitman (1949), which has been the dominant method of comparison in econometrics and statistics; see Engle (1984), Newey and McFadden (1994), van der Vaart (1998), and Lehmann and Romano (2005) for a discussion. Also considered has been comparing tests when $\alpha \to 0$ as $n \to \infty$. This leads to the comparison of Bahadur (1960); see also Bahadur (1967). For a broad overview of these and other relative efficiency measures, see Serfling (2009). In this paper we consider comparisons when $\beta \to 0$ because α and μ are fixed. First proposed in Hodges and Lehmann (1956), it turns out that for consistent tests we have that

$$\lim_{n\to\infty} \frac{h(n)}{n} = \lim_{n\to\infty} \frac{\log P(\phi_2 \text{ rejects } H_0)}{\log P(\phi_1 \text{ rejects } H_0)} =: e(\phi_1, \phi_2) \tag{3}$$

where the probabilities are evaluated for $\mu \neq \mu_0$ is fixed. This implies that when the limiting ratio of log-probabilities exists, we can compare two tests asymptotically in a Neyman-Pearson sense without requiring $\mu \to \mu_0$ as $n \to \infty$. Clearly, if $e(\phi_1, \phi_2) > 1$, we prefer ϕ_2, and visa versa when $e(\phi_1, \phi_2) < 1$. In general, (3) is computed by evaluating:

$$\lim_{n\to\infty} -\frac{1}{n} \log P(\phi_i \text{ rejects } H_0) \tag{4}$$

For consistent tests, if this limit exists it must be non-negative. It also turns out that the limit in (4), is bounded above by the Kullback-Leibler divergence $K(\mathcal{P}_0, P_1)$ between the set of null distributions $\mathcal{P}_0$ and the true distribution P_1, where we define:

$$K(\mathcal{P}_0, P_1) = \inf_{P\in\mathcal{P}_0} \int_{\mathbb{R}} \log\left(\frac{dP}{dP_1}\right) dP.$$

Here, $K(\mathcal{P}_0, P_1) = \infty$ if there does not exist any $P \in \mathcal{P}_0$ such that P is absolutely continuous with respect to P_1. Under the assumption that the sample space is compact, it has been shown in Canay and Otsu (2012) that (2) lead to efficiency tests in that (4) is equal to $K(\mathcal{P}_0, P_1)$.

Our main result is that the t-test achieves the lower-bound of 0 when the observations have heavy-tails. By heavy-tails, in this paper we mean that for all $\epsilon > 0$, we have that:

$$\sup_{s\in(-\epsilon,\epsilon)} \mathbb{E}\, e^{sX_i} = \infty \tag{5}$$

Equivalently, the moment generating function diverges in any neighborhood of zero. We allow for one tail of the distribution to be light, in that the supremum can be finite if we take the supremum over $s \in (-\epsilon, 0)$ or $(0, \epsilon)$, but at least one tail will be heavy in what follows.

Before stating our main result, we will discuss the necessary assumptions in turn. Our first assumption is standard for validity of the t-test.

Assumption 2.1. *We assume the X_i are i.i.d. with mean μ, variance $\sigma^2 < \infty$, with common distribution function $F(x)$.*

This first assumption leads to the t-test providing for valid inference, and in fact achieving the semiparametric efficiency bound, in the sense of Pitman efficiency, asymptotically; see Levit (1976), Newey (1990), and Newey and McFadden (1994).

This next assumption implies (5), but is slightly stronger, and provides us the main regularity conditions we need for the results.

Assumption 2.2. *We assume that for all* $\lambda > 0$*:*

$$\lim_{x\to\infty} \frac{\log P(|X_i| > \lambda x)}{\log P(|X_i| > x)} = 1 \tag{6}$$

We also assume that the following tail-balance conditions holds:

$$\lim_{x\to\infty} \frac{P(X_i > x)}{P(|X_i| > x)} = p \in [0, 1] \tag{7}$$

Furthermore, we assume the heavier tail is subexponential: for any $y \in (-\infty, \infty)$*,*

$$\lim_{x\to\infty} \frac{P(X_i > x + y)}{P(X_i > x)} = 1,\ p \in (0, 1]$$
$$\lim_{x\to\infty} \frac{P(X_i < -x + y)}{P(X_i < -x)} = 1,\ p \in [0, 1)$$

In Assumption 2.2, we assume that the cumulative hazard function of X_i or $-X_i$ is slowly-varying. A function L is slowly-varying if for all $\lambda > 0$, $\lim_{x\to\infty} L(\lambda x)/L(x) = 1$. Here, (6) implies (5), however it includes some distributions, such as the lognormal distribution, with all finite moments. Also included are distributions with approximately polynomial tails, such as student-t and Pareto distributions. Distributions with two thin-tails such as the Gaussian or exponential distribution are not included, although we allow for one of the two tails to be light-tailed. Our proof methods focus on the behavior of the maximum absolute order statistic, $\max_i |X_i|$. The tail-balance condition in (7) rules out scenarios where we cannot determine the behavior of $\max_i |X_i|$ because the tails of F are fluctuating between radically different behavior. An example of what is ruled out are distributions which interpolate between different rates of polynomial decrease, no matter how far out in the tail we look. The assumption of subexponential tails is standard in the literature on large deviations of heavy-tailed sums. In Cline and Hsing (1989), lower-level technical conditions are used to rule out heavy-tailed random variables which are not subexponential, however for our purposes it is both simpler to state subexponentiality as an assumption, and easier for the reader to check which distributions satisfy this assumption. All random variables with regularly-varying tails satisfy this assumption, as well as lognormal random variables. There are subexponential random variables which do not satisfy (6), such as Weibull random variables with infinite moment generating function. Likewise, there are random variables which satisfy (6) which are not subexponential; see e.g. Foss et al. (2011), Section 3.7. It would be of interest to derive our results for all subexponential random variables, however this is outside the scope of the present paper.

Assumption 2.3. *The absolute observations $|X_i|$ are in the maximum domain of attraction of the Frechét distribution or Gumbel distribution. When the $|X_i|$ are in the maximum domain of attraction of the Gumbel distribution, we also require the Von-Mises condition is satisfied. We assume that F has continuously differentiable density f, and we define:*

$$h_+(x) := \frac{f(x)}{1 - F(x)}, \quad h_-(x) := \frac{f(-x)}{F(-x)}$$

These are the hazard functions for X_i and $-X_i$ respectively. For $|X_i|$ in the domain of attraction of the Gumbel distribution, we require that the appropriate von Mises condition(s) is met:

$$p < 1 : \lim_{x\to\infty} \frac{d}{dt}\frac{1}{h_+(t)}\bigg|_{t=x} = 0, \qquad p > 0 : \lim_{x\to\infty} \frac{d}{dt}\frac{1}{h_-(t)}\bigg|_{t=x} = 0 \tag{8}$$

Furthermore, we require that $h_+(t)\sqrt{t} \to 0$ and $h_-(t)\sqrt{t} \to 0$ when $p \in (0, 1]$ or $p \in [0, 1)$ respectively.

These assumptions are natural in the following sense: in the case of heavy-tailed data, Type-II errors are generated by the maximum absolute order statistics. Thus, we require some regularity in the behavior of these extrema to be able to formulate our results. The Von Mises condition is satisfied, for example, by lognormal-type random variables, i.e. where for some $\eta > 0$,

$$\lim_{x\to\infty} \frac{-\log P(|X_i| > x)}{\log^\eta(x)} = 1$$

In the case of $\max_i |X_i|$ being in the domain of attraction of a Frechét random variable, necessary and sufficient conditions are that there exists $\gamma > 0$ and a slowly-varying function L such that:

$$\lim_{x\to\infty} \frac{P(|X_i| > x)}{x^{-\gamma}L(x)} = 1 \tag{9}$$

This implies that $1 - F(x) + F(-x)$ is regularly-varying; see Bingham et al. (1989) for an exhaustive treatment of regularly varying functions, including their application in probability. Examples satisfying (9) include all student-t and Pareto random variables. Recall that in the case of the t-test, we choose a critical value C_α so that under the null-hypothesis $P(T_n > C_\alpha) \to \alpha$. We place some restrictions on the range of C_α, and therefore α, which we allow.

Assumption 2.4. *We require $C_\alpha > 1$.*[1]

Assumption 2.4 restricts which the range of α we consider for the asymptotic size of the test. This requirement is a unifying requirement. Results can be derived for $C_\alpha \leq 1$, however they cannot be stated as succinctly. When $C_\alpha > 1$, it turns out that observations from both tails contribute to the type-II error rate when the tail balance parameter satisfies $p \in (0, 1)$. This is different from the case where σ is known: in this case, especially small values of Z_n that lead to type-II errors only come from the left tail of the distribution. When the variance is being estimated, large observations from the right tail contribute to shrinking the test statistic and producing type-II errors.

Theorem 1. *Recall that T_n is the typical t-test statistic given by (2). When the distribution of the X_i satisfies Assumptions 2.1-2.4, we have that whenever $\mathbb{E} X_i = \mu > \mu_0$, then for $R(n) = -\log nP(|X_i| > n)$, we have that:*

$$\lim_{n\to\infty} \frac{-\log P(T_n < C_\alpha)}{R(n)} = 1 \tag{10}$$

Remark 1. One immediate implication of Theorem 1 is that the convergence rate of the type-II error is slower than exponential, that is to say:

$$\lim_{n\to\infty} -\frac{1}{n} \log P(T_n < C_\alpha) = 0$$

This implies that the t-test achieves the lower bound among consistent tests in (4). Thus, the t-test is far from efficient in a Hodges-Lehmann sense. The proof of (10) involves constructing a lower and upper bound for $\log P(T_n < C_\alpha)$. The lower bound is valid under weaker conditions than those used in the theorem. This implies that the conclusion $\lim_{n\to\infty} n^{-1} \log P(T_n < C_\alpha) = 0$ holds for a much broader class of distributions than those we consider here.

Remark 2. When the observations are in the Frechét domain of attraction, we can also choose $R(n) = (\gamma - 1)\log n$. This will inform our strategy for constructing a robust test statistic later. If we can construct a test ϕ such that for some sequence a_n:

$$\lim_{n\to\infty} \frac{-\log P(\phi \text{ rejects } H_0)}{a_n} \geq 0$$

and $R(n)/a_n \to 0$, then the relative efficiency of ϕ to the t-test will be ∞.

[1] Implying we the level for testing (1) is at most 0.158

Remark 3. The proof is based on techniques from Cline and Hsing (1989), Mikosch and Nagaev (1998), and Lehtomaa (2017). Those papers develop large deviation results for sums of i.i.d. random variables with heavy tails, including regularly varying tails. The essential idea is that when the tails of the X_i are heavy, asymptotically large-deviation probabilities of a sum are equal to large deviations of the maximum. The conditions used here are strictly stronger than those used the probability literature, and therefore there is some room to gain generality in the theorem as it stands now.

Theorem 1 is a negative result, in that it highlights an inefficiency of the t-test when dealing with heavy-tailed data. It is also instructive in that it gives an asymptotic justification for robust methods. Later in the paper we will discuss a couple of robust procedures that can improve the efficiency of hypothesis tests in this heavy-tailed setting.

2.2 Comparison With Local Asymptotic Power

The most common approximation of the asymptotic power of tests is local asymptotic power. Under Assumption 2.1 we have that

$$T_n \Rightarrow \mathcal{N}(0,1)$$

under the null hypothesis. To compute relative efficiency in the sense of Pitman, we introduce a sequence of alternatives μ_n, such that $\mathbb{E}\, X_i = \mu_n = \mu_0 + \delta/\sqrt{n}$. Under this sequence of alternatives, we have that the test statistics converge to shifted normal random variables:

$$T_n \Rightarrow \mathcal{N}(\delta/\sigma, 1) \tag{11}$$

An implication of (11) is that under a local asymptotic power comparison, the power properties of each test are the same for all distributions with the same variance σ^2. Furthermore, the t-test is semiparametric minimax efficient, as discussed previously. By contrast, in Theorem 1, not only does the asymptotic power depend on the variance σ, it also depends on the tail properties of the distribution, namely $\log nP(|X_i| > n)$. In the case where $P(|X_i| > x)$ is regularly varying, for example, the tail index γ plays a role. Thus, our results provide for a finer distinction relative to local asymptotics in this setting.

2.3 Bahadur Relative Efficiency

Another notion of relative efficiency, due to Bahadur (1960), is to compare the rate at which the p-values of a test converge to 0 under a fixed alternative. If we denote the sequence of distribution functions of a test statistic W_n under the null as G_n, where we reject when W_n is large, then the sequence of p-values is given by:

$$1 - G_n(W_n). \tag{12}$$

In He and Shao (1996), it is shown that the p-values converge to zero under a fixed alternative at an exponential rate, and in fact under Assumption 2.1:

$$\lim_{n\to\infty} -\frac{1}{n}\log(1 - G_n(T_n)) = -\log\left(\sup_{c\geq 0}\inf_{t\geq 0}\mathbb{E}\exp\left\{2tcX_i - \frac{\Delta}{\sigma}\frac{t(c^2+X_i^2)}{\sqrt{1+\Delta^2/\sigma^2}}\right\}\right). \tag{13}$$

Note that since the second term inside the exponent is positive and quadratic in X_i, t will always be chosen larger than 0. Therefore, this result implies that for the t-test statistic, p-values converge to 0 at an exponential rate. Notice that this suggests in heavy-tailed cases that T_n should be used rather than Z_n. Suppose that $P(-X_i > x)$ is regularly-varying with index γ, i.e. the left-tail of the distribution of the X_i is approximately polynomial. For Z_n, under Assumptions 2.1-2.4, by Theorem 3.3 in Cline and Hsing (1989),

$$\lim_{n\to\infty} -\frac{1}{\log n}\log\left(1 - G_n\left(Z_n\right)\right) = \gamma - 1 \tag{14}$$

Thus, when using Bahadur relative efficiency to compare Z_n and T_n, T_n appears to have additional robustness of well-controlled p-values. Essentially, (13) says that under the null hypothesis, when n is large the test statistic has thin tails. Thus, large p-values disappear at an exponential rate. This does not, however, say anything about the probability that a given p-value is under the desired significance level. Since test statistics and p-values are in 1-to-1 correspondence, an interpretation of Theorem 1 is that when X_i has heavy tails, the probability a given p-value exceeds the desired significance level is disappearing at a polynomial rate. In this sense, Bahadur relative efficiency can be framed as a comparison based on the behavior of the maximum p-value, and Hodges-Lehmann relative efficiency is based on an expectation over the p-values. This difference leads to Bahadur relative efficiency implying Z_n has worse properties than T_n. On the other hand, consider the case when the left tail of X_i is thin, or there exists a $s > 0$ such that $\mathbb{E}\, e^{-sX_i} < \infty$. In this case, it turns out that

$$\lim_{n\to\infty}\frac{\log P(Z_n < C_\alpha)}{\log P(T_n < C_\alpha)} = 0$$

even when the right tail of the distribution of the X_i is heavy. In the latter case, T_n will still have slower than exponential rates of convergence. If the left tail is heavy, the two tests will be equivalent. Thus, in a Hodges-Lehmann setting, we would weakly prefer the test with known variance when the variance is in fact known.

3 A New Robust Testing Procedure

3.1 Naive robust tests via subsampling

We now propose a new testing procedure designed to address the inefficiencies of the t-test. We first motivate the testing procedure then discuss the construction of the test statistic. There has been recent work in the machine learning literature on obtaining concentration guarantees for estimators. Specifically, this literature has focused on constructing estimators $\hat{\theta}$ of a parameter θ such that:

$$\log P(\|\hat{\theta} - \theta\| > \delta) < -g(\delta)$$

for certain choices of δ and convex g. See Minsker (2019), Sun et al. (2020), and Mathieu and Minsker (2021) for examples. We can rearrange (10) to see how this type of thinking could help us. Under a fixed alternative, $n^{-1/2}T_n \xrightarrow{P} \Delta/\sigma$. Consider then that we can re-write our type-II error as:

$$P(T_n < C_\alpha) = P\left(\frac{1}{\sqrt{n}}T_n - \frac{\Delta}{\sigma} < -\frac{\Delta}{\sigma} + \frac{1}{\sqrt{n}}C_\alpha\right)$$

Thus, we can interpret the shortcoming of the t-test as a failure of the test statistic to be sufficiently concentrated around the non-centrality parameter Δ/σ in large samples.

The idea we use to construct a robust test is to use sample splitting and robust recombination to generate test statistics with thinner tails, and more concentration around the non-centrality parameter, under the alternative. Start by splitting the sample into k groups of size m, so that $km = n$. On each sub-sample, construct the t-statistic $T_{j,m}$, $j = 1, \ldots, k$. Let ρ be a symmetric, convex function which is minimized at 0. Consider the statistic:

$$\tilde{\tau} = \arg\min_t \sum_{j=1}^{k} \rho(T_{j,m} - \sqrt{m}t)$$

When $\rho(t) = \frac{1}{2}t^2$, $\sqrt{m}\tilde{\tau} = \frac{1}{k}\sum_{j=1}^{k} T_{j,m}$. When $\rho(t) = |t|$, $\sqrt{m}\tilde{\tau} = \text{Median}(T_{j,m})$. In general, we might expect that:

$$\sqrt{n}\tilde{\tau} = \left(\frac{1}{k}\sum_{j=1}^{k} \rho''(T_{j,m})\right)^{-1} \frac{1}{\sqrt{k}}\sum_{j=1}^{k} \rho'(T_{j,m}) + o_P(1)$$

If we choose ρ such that ρ' is bounded, then we now have a test-statistic with a bounded influence function. Unfortunately, this procedure will tend to be inefficient. Generally, the asymptotic variance of $\sqrt{n}\tilde{\tau}$ will be larger than 1. This leads us to seek a procedure that retains efficiency, while still leading to a robust testing procedure.

3.2 Efficient tests based on subsampling

The results of the previous section are unsatisfactory since the proposed procedure is not permutation invariant. Conditional on an observed data set, different sample splits could lead to different empirical conclusions. A solution, which also turns out to improve efficiency, is to extend the results of the previous section to permutation-invariant analogues. To start, we define $\mathcal{A}_{m,n}$ as the set of all subsets of $\{1, \ldots, n\}$ of size m. An element $A_j \in \mathcal{A}_{n,m}$ is therefore:

$$A_j = \{j_1, \ldots, j_m\},\ j_i \neq j_k,\ i \neq k$$

Given our sample of i.i.d. observations, we can compute the t-statistic on each subsample: for each A_j, we use $\{X_{j_k}\}_{j_k \in A_j}$ to compute a t-statistic, denoted $T_m(A_j)$. Let ρ be a convex function. We will use $\sqrt{n}\hat{\tau}_{n,m}$ as our new test-statistic, where

$$\hat{\tau}_{n,m} := \arg\min_t \frac{1}{\binom{n}{m}} \sum_{A_j \in \mathcal{A}_{n,m}} \rho\left(T_m(A_j) - \sqrt{m}t\right). \tag{15}$$

We will generally abbreviate to $T_{j,m} = T_m(A_j)$ where there is no confusion. Our first set of additional assumptions concern the smoothness of ρ:

Assumption 3.1. *We assume that ρ is three-times continuously differentiable, $\|\rho''\|_\infty, \|\rho'''\|_\infty < \infty$, $\rho''(x) \geq 0$ for all x, and $\mathbb{E}\,\rho''(Z) > 0$, where $Z \sim \mathcal{N}(0, 1)$.*

The boundedness of the derivatives of ρ ensures that the objective function in (15) is well-approximated by a quadratic function in large samples. The additional conditions on the second derivative are necessary for consistent inference for the mean, and essentially guarantee that the objective function is sufficiently convex near 0.

The next assumption considers a certain type of compatibility between ρ and the sequence of t-statistics.

Assumption 3.2. *We assume that the sequence $\rho'(T_{j,m})^2$ is uniformly-integrable.*

This assumption will trivially be met when ρ' is uniformly bounded. Simple choices include the smoothed-Huber function, $\rho(x) = \sqrt{1+x^2} - 1$ and log-cosh $\rho(x) = \log(\cosh(x))$. When ρ' is not bounded, then we need some knowledge of when the moments of $\mathbb{E}\,\rho'(T_{j,m})$ exist. One set of simple sufficient conditions is $\mathbb{E}\,|T_{j,m}|^{2+\delta} < \infty$ for all m and $|\rho'(t)|/|t| \to 0$ as $t \to \pm\infty$. See Jonsson (2011) for sufficient conditions for the existence of moments of the t-statistic. One such case we briefly mention here is that when X_i has a density function $f(x)$ such that $f(x)$ is eventually monotone: there exist M_-, M_+ such that for all $x < y < M_-$, $f(x) < f(y)$, and for all $x > y > M_+$, $f(x) < f(y)$.

Let $T_{j,m}(X_{-1}, x)$ denote the t-statistic formed from $X_2, \dots, X_n$ and x. We also note that the uniform integrability condition in Assumption 3.2 and the smoothness assumption in Assumption 3.1 imply, for the t-statistic, that:

$$\mathbb{E}(\sqrt{m}(\mathbb{E}[\rho'(T_{j,m}(X_{-1}, X_1)) - \rho'(T'_{j,m}(X_{-1}, X'_1))|X_1] - \mathbb{E}\,\rho''(T_{j,m})) \to 0 \tag{16}$$

where X'_1 is independent of X_1, but identically distributed. Thus, our results can be expected to hold for a broader class of statistics that satisfy the uniform integrability condition for a suitable choice of ρ. The additional part to show is that the contribution of an individual observation to the distribution of the statistic disappears at a sufficiently fast rate in sample size. This property holds for the t-statistic, and can be expected to hold for other statistics which are asymptotically linear and satisfy a Central Limit Theorem.

Our final assumption relates to the choice of m.

Assumption 3.3. *We assume that $m/n \to 0$ and $m \to \infty$.*

This assumption starts to make clear the sense in which m/n serves the role of bandwidth. If m does not go to ∞, then we cannot guarantee that our procedure will be consistent for inference on the mean when the data are skewed. $m/n \to 0$ plays an important role both in the quadratic approximation to the objective function in (15) as well as the central limit theorem we must apply. Lastly, we define:

$$\tau_{n,m} := \frac{1}{\sqrt{m}} \frac{\mathbb{E}\,\rho'(T_{j,m})}{\mathbb{E}\,\rho''(T_{j,m})}$$

We are now ready to state the main result of this section.

Theorem 2. *Under Let Assumptions 2.1, 3.1, and 3.2, let $\hat{\tau}_{n,m}$ be defined as in (15). Fur-*

thermore, let $\mathbb{E}\, X_i = \mu_0 + \delta/\sqrt{n}$*. Then, we have that:*

$$\sqrt{n}(\hat{\tau}_{n,m} - \tau_{n,m}) \Rightarrow \mathcal{N}(\delta/\sigma, 1) \tag{17}$$

The theorem states $\sqrt{n}\hat{\tau}_{n,m}$ has the same asymptotic variance as the t-statistic, but it will generally be biased. Apart from the bias term, observe that the asymptotic distribution of the test statistic does not depend on ρ. Namely, when the bias is asymptotically negligible, the choice of ρ has no impact on the first-order properties of the test statistic. When is $\tau_{n,m}$ small? If ρ' is an odd function and the X_i are symmetrically distributed then $\tau_{n,m} = 0$ for all n, m. More generally, if $\|\rho'\|_\infty = C < \infty$, then we can bound $\tau_{n,m}$ using the convergence rate of the t-statistics to normal random variables

$$\mathbb{E}\, \|\rho'(T_{j,m})\| = \mathbb{E}\, \|\rho'(T_{j,m}) - \rho'(Z)\| \leq 2C \sup_{x\in\mathbb{R}} |P(T_{j,m} < x) - \Phi(x)|.$$

Thus, we can bound the distance $\tau_{n,m}$ is from 0 by Berry-Esseen type bounds. Namely, if there exists $\epsilon \in (0, 1]$ such that $\mathbb{E}\, |X_i|^{2+\epsilon} < \infty$, then $\sqrt{n}\tau_{n,m} = O(\sqrt{n}m^{-(\epsilon+1)/2})$. Thus, if we choose m such that $n/m^{1+\epsilon} \to 0$, then $\sqrt{n}\tau_{n,m} \to 0$. We have shown that using $\sqrt{n}\hat{\tau}_{n,m}$ preserves the first-order local efficiency properties of the t-statistic. What about the large deviation properties? The next result says that constructing consistent tests which avoid worst-case rates of convergence for the Type-II error rate is straightforward in this setting.

Theorem 3. *In addition to the assumptions for Theorem 2, assume that* ρ *is symmetric* $\|\rho'\|_\infty < \infty$*, and* $\mathbb{E}\, X_i = \mu_0 + \Delta$ *rather than the local-alternative. Then:*

$$\lim_{n\to\infty} -\frac{m}{n} \log P(\sqrt{n}\hat{\tau} < C_\alpha) \geq \frac{1}{4} \tag{18}$$

An important consequence of Theorem 3 is that for all of the heavy-tailed distributions considered in this paper, if $m = an^\zeta$ for some ζ,

$$\lim_{n\to\infty} \frac{\log P(\sqrt{n}\hat{\tau} < C_\alpha)}{\log P(T_n < C_\alpha)} = \infty$$

Thus, the new testing procedure is also more efficient than the t-test under Hodges-Lehmann asymptotic relative efficiency.

4 Simulations

We provide simulation results calibrated to arithmetic returns from the SPDR S&P 500 ETF Trust (SPY), as in Shephard (2020). We simulate data from the following DGP:

$$\begin{aligned} y_i &= (z_i - \psi)\beta_1 + \varepsilon_i \\ z_i &= \psi + V\sigma_z \\ \varepsilon_i &\sim \mathcal{N}(0, (1 + |z_i - \psi|^\zeta)C^2) \end{aligned} \tag{19}$$

We choose three different distributions for V:

$$\sqrt{\frac{\nu}{\nu - 2}} V_\nu \sim t_\nu, \quad \nu \in \{\nu_l, \nu_h\}$$
$$\exp\{-v/4\}V \sim \mathcal{N}(0,\, \eta_i^{-1}), \quad \log \eta_i \sim \mathcal{N}(0, v),\, v \in \{v_l,\, v_h\}$$

where t_ν is a student-t distribution with ν degrees of freedom. Here, z_i are calibrated to match the weekly returns. We use data from 19th November, 2018, through 18th November, 2022. The first two years of data exhibit more volatility due to inclusion of the start of the global pandemic in spring of 2022. We use the first two years to estimate ν_h and v_h for our "higher volatility" setup. We used the full two years to calibrate ν_l and v_l. We set $\nu_h = 2.219$, $\nu_l = 3.352$, $v_l = 0.933$, and $v_h = 1.663$. We then set σ_Z so that the standard deviation matches the sample standard deviation of the returns during the relevant time period, and ψ is likewise the sample mean for each time period. We assume ψ is known, so that we can perform a heteroskedasticity-robust test of the null $H_0 : \beta_1 = 1$ using the t-statistic formed from the observations $X_i = (z_i - \psi)(y_i - (z_i - \psi)\beta_0)$. Unlike Shephard (2020), we include conditional heteroskedasticity. We do this so that the t-test is clearly a reasonable thing to do here, rather than a more efficient procedure that leverages homoskedasticity. The form of (19) is motivated so that for our choice of ζ, T_n provides asymptotically valid inference under the null, and we choose ζ to be 0.12 in the high-volatility case when z_i are student-t, and set $\zeta = 0.5$ in the lower-volatility setting. For the lognormal mixtures, we set $\zeta = 1.0$. C is chosen so that $\mathbb{E}\,\varepsilon_i^2 = 4$, as in Shephard (2020). We set the null value of β as $\beta_{1,0} = 1$. Our simulations will display the power of the relevant tests as n increases, for a fixed alternative. We expect our new procedure to improve upon the standard t-test, so we chose $\beta_{1,1}$ for each simulation setting so that the power of our new test was 0.99 when $n = 250$.

To investigate the finite-sample power properties of our new testing procedure, we performed a simple one-sided test for the mean with samples ranging from 50 to 500 observations. We also chose different growth paths for m of the form $m = an^{\epsilon}$ for the robust testing procedure. The choices for ϵ were $(0.9, 0.5, 0.2, 0.1)$. a was chosen so that the smallest m would be 30, the classical rule of thumb used in introductory statistics for the Central Limit Theorem to hold. For the choice of ρ we use the smoothed-Huber function:

$$\rho(x) = \sqrt{1 + x^2} - 1$$

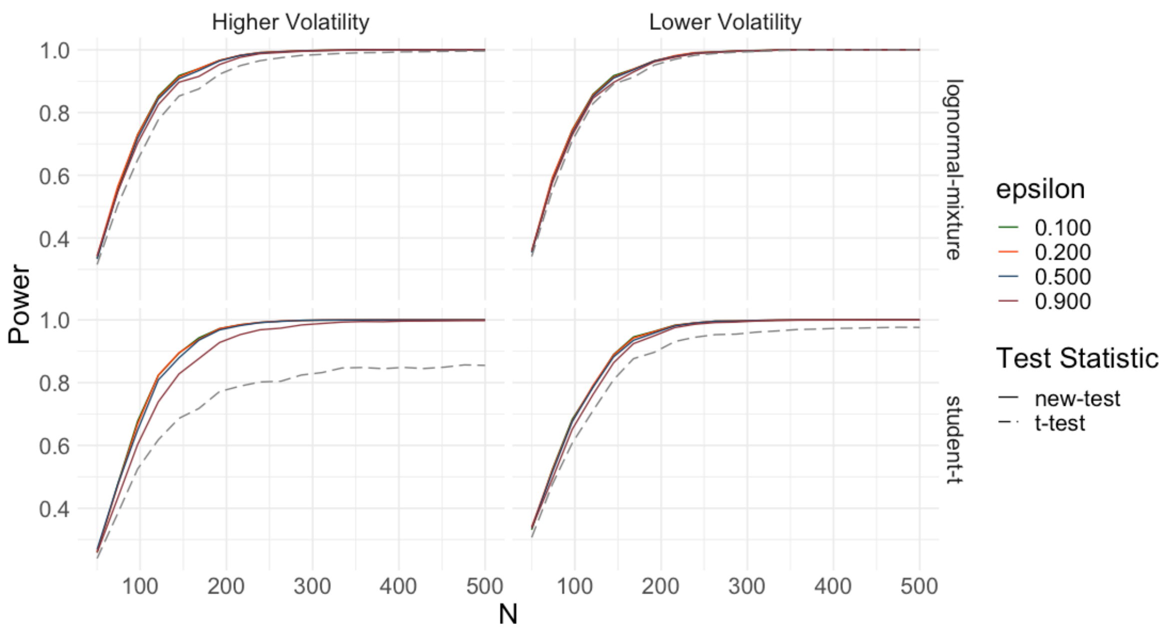


Figure 1: Comparison of New Test with t-test
This figure plots the power when testing $\beta_1 = 1$ against a one-sided alternative $\beta_1 > 1$. The power curves have been (point-wise in sample size) size-corrected. The dashed line corresponds to the classic t-test, while the other four lines correspond to the new testing procedure, with different choices for how the size of subsamples m grows with sample size.

In Figure 1, we plot power curves as a function of sample size. All power calculations are performed after correcting the critical values so that the size of all tests is 0.01. Notice that the sample size required by the t-test to achieve a particular power level is uniformly larger than the sample size required by the new robust procedure. Furthermore, the new test gets arbitrarily close to 1 for all data generating processes considered, as sample size increases, whereas the power curve for the traditional t-test flattens out considerably as the sample size gets large, especially when the data are student-t. This supports the claim in Theorem 3 that the new testing procedure leads to better Type-II error properties. Also notice that the choice of ϵ does not matter much qualitatively here, with the possible exception of a small power loss when the tails are heaviest.

5 Conclusion

This paper provides new results on the efficiency of t-tests when the data have heavy tails. These results are the first to demonstrate the inefficiency of the t-test with heavy-tailed data in an asymptotic setting. Our results complement recent work studying the performance of test statistics under the null hypothesis when the data exhibit heavy tails.

It would be desirable to extend the main results to cases in which the observations have "nearly" exponential tails, such as Weibull-type tails. Other useful extensions include extending these results to classical Wald statistics and more generally to GMM-type statistics. It would also be interested to consider how to use Theorem 10 to conduct power analysis; currently, the limiting type-II error rate diverges as the alternative approaches the null, implying that higher order terms and knowledge of the function R might be useful in practice. It is of course an open question if it is possible to recover Hodges-Lehmann efficiency for semiparametric inference on the mean in these heavy-tailed settings.

References

Bahadur, R. R. (1960): "Stochastic Comparison of Tests," *The Annals of Mathematical Statistics*, 31, 276–295.

——— (1967): "Rates of Convergence of Estimates and Test Statistics," *The Annals of Mathematical Statistics*, 38, 303–324.

Bingham, N. H., C. M. Goldie, J. L. Teugels, and J. Teugels (1989): *Regular Variation*, 27, Cambridge University Press.

Brosset, F., T. Klein, A. Lagnoux, and P. Petit (2022): "Large Deviations at the Transition for Sums of Weibull-like Random Variables," in *Séminaire de Probabilités LI*, Springer, 239–257.

Brownlees, C., E. Joly, and G. Lugosi (2015): "Empirical Risk Minimization for Heavy-Tailed Losses," *The Annals of Statistics*, 43, 2507–2536.

Canay, I. A. and T. Otsu (2012): "Hodges–Lehmann Optimality for Testing Moment Conditions," *Journal of Econometrics*, 171, 45–53.

Cline, D. B. and T. Hsing (1989): "Large Deviation Probabilities for Sums of Random Variables with Heavy or Subexponential Tails," Tech. rep., Texas A&M University.

Dembo, A. and O. Zeitouni (2009): *Large Deviations Techniques and Applications*, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg.

Efron, B. (1969): "Student's t-Test under Symmetry Conditions," *Journal of the American Statistical Association*, 64, 1278–1302.

Engle, R. F. (1984): "Chapter 13 Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics*, Elsevier, vol. 2, 775–826.

Foss, S., D. Korshunov, S. Zachary, et al. (2011): *An Introduction to Heavy-Tailed and Subexponential Distributions*, vol. 6, Springer.

Giné, E., F. Götze, and D. M. Mason (1997): "When Is the Student T-Statistic Asymptotically Standard Normal?" *The Annals of Probability*, 25, 1514–1531.

Hall, P. (1992): "On the Removal of Skewness by Transformation," *Journal of the Royal Statistical Society: Series B (Methodological)*, 54, 221–228.

He, X. and Q.-m. Shao (1996): "Bahadur Efficiency and Robustness of Studentized Score Tests," *Annals of the Institute of Statistical Mathematics*, 48, 295–314.

Hjort, N. L. and D. Pollard (2011): "Asymptotics for Minimisers of Convex Processes," *arXiv preprint arXiv:1107.3806*.

Hodges, J. and E. Lehmann (1956): "The Efficiency of Some Nonparametric Competitors of the T-Test," *Annals of Mathematical Statistics*, 27, 324–335.

Jonsson, F. (2011): "On the Heavy-Tailedness of Student's t -Statistic," *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 17, 276–289.

Lee, A. J. (2019): *U-Statistics: Theory and Practice*, Routledge.

Lehmann, E. L. and J. P. Romano (2005): *Testing Statistical Hypotheses*, vol. 3, Springer.

Lehtomaa, J. (2017): "Large Deviations of Means of Heavy-Tailed Random Variables with Finite Moments of All Orders," *Journal of Applied Probability*, 54, 66–81.

Levit, B. Y. (1976): "On the Efficiency of a Class of Non-Parametric Estimates," *Theory of Probability & Its Applications*, 20, 723–740.

Mathieu, T. and S. Minsker (2021): "Excess Risk Bounds in Robust Empirical Risk Minimization," *Information and Inference: A Journal of the IMA*, 10, 1423–1490.

Mikosch, T. and A. V. Nagaev (1998): "Large Deviations of Heavy-Tailed Sums with Applications in Insurance," *Extremes. Statistical Theory and Applications in Science, Engineering and Economics*, 1, 81–110.

Minsker, S. (2019): "Distributed Statistical Estimation and Rates of Convergence in Normal Approximation," *Electronic Journal of Statistics*, 13, 5213–5252.

——— (2022): "U-Statistics of Growing Order and Sub-Gaussian Mean Estimators with Sharp Constants," *arXiv preprint arXiv:2202.11842*.

Müller, U. K. (2019): "Refining the Central Limit Theorem Approximation via Extreme Value Theory," *Statistics & Probability Letters*, 155, 108564.

——— (2020): "A More Robust T-Test," *arXiv:2007.07065 [econ.EM]*.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, 4, 2111–2245.

Peng, W., T. Coleman, and L. Mentch (2019): "Asymptotic Distributions and Rates of Convergence for Random Forests via Generalized U-Statistics," *arXiv preprint arXiv:1905.10651*.

Pitman, E. J. (1949): "Notes on Non-Parametric Statistical Inference," Tech. rep., North Carolina State University. Dept. of Statistics.

Serfling, R. J. (2009): *Approximation Theorems of Mathematical Statistics*, vol. 162, John Wiley & Sons.

Shao, Q.-M. (1997): "Self-Normalized Large Deviations," *The Annals of Probability*, 285–328.

——— (1999): "A Cramér Type Large Deviation Result for Student's t-Statistic," *Journal of Theoretical Probability*, 12, 385–398.

Shephard, N. (2020): "An Estimator for Predictive Regression : Reliable Inference for Financial Economics," Tech. rep., Harvard University.

Student (1908): "The Probable Error of a Mean," *Biometrika*, 1–25.

Sun, Q., W.-X. Zhou, and J. Fan (2020): "Adaptive Huber Regression," *Journal of the American Statistical Association*, 115, 254–265.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge University Press.

Victor, H., D. la Peña, T. L. Lai, and Q.-M. Shao (2009): *Self-Normalized Processes: Limit Theory and Statistical Applications*, vol. 204, Springer.

Young, A. (2021): "Leverage, Heteroskedasticity and Instrumental Variables in Practical Application"."," Tech. rep., London School of Economics.

A Proofs of Main Results

A.1 Proof of Theorem 1

In this subsection we provide a proof of (10). The result will follow immediately from Lemmas 1-5. The application to the testing problem is new, however we should note that many of the proof techniques closely follow those used in Cline and Hsing (1989), Mikosch and Nagaev (1998), Brosset et al. (2022), and assuredly other papers in the literature on large deviations of sums of heavy-tailed random variables. We begin with an equivalency relationship that has been previously observed, in e.g. Shao (1997) and Victor et al. (2009):

$$[T_n < C_\alpha] = [S_n < c_n V_n] \tag{20}$$

where $c_n \to C_\alpha$ monotonically from above as $n \to \infty$, and $S_n = \sum_{i=1}^n Z_i$, $V_n^2 = \sum_{i=1}^n Z_i^2$, and $Z_i = (X_i - \mu_0)/\sigma$. It is also helpful to denote $Y_i = Z_i - \nu$, so that $\mathbb{E}\, Y_i = 0$ and $\mathrm{Var}(Y_i) = 1$, and $Q_n = \sum_{i=1}^n Y_i$. It is also helpful to denote $S_{n,i}$ and $Q_{n,i}$ as $S_n - Z_i$ and $Q_n - Y_i$ respectively. Dependence of c_n on n is not important for our purposes, and thus we will omit the subscript. Recall that since $R(n) = -\log nP(|X_i| > n)$, and $\mathbb{E}\,|X_i| < \infty$, we have that $nP(|X_i| > n) \to 0$, and thus for n large enough, $R(n) > 0$.

We begin by providing a lower-bound for the type-II error rate $P(T_n < C_\alpha)$. The following lemma summarizes our results:

Lemma 1. *Assume the conditions of Theorem 1 hold. Then, we have that:*

$$\lim_{n\to\infty} \frac{\log P(T_n < C_\alpha)}{R(n)} \geq -1$$

Proof. We split things into two cases. We begin by assuming the left tail of X_i is the heavier tail, which is equivalent to saying $p = 0$, in the notation of Assumption 2.2.

(Case 1) In this case $p = 0$. We have that for any $K > 0$,

$$\begin{aligned}
P(T_n < C_\alpha) &= P(S_n < cV_n) \geq P(S_n < cV_n,\, \min_i Z_i < -Kn) \\
(2-\infty \text{ norm inequality}) &\geq P(S_n < c\max_i |Z_i|,\, \min_i Z_i < -Kn) \\
&\geq P(S_n < c(-\min_i Z_i),\, -\min_i Z_i > Kn) \\
&\geq P(S_n < cKn, \min_i Z_i < -Kn) \\
(\text{Bonferroni inequality}) &\geq \sum_{i=1}^{n} P(S_n < cKn, Z_i < -Kn) \\
&\quad - \sum_{i \neq j} P(S_n < cKn,\, Z_i < -Kn,\, Z_j < -Kn) \\
&\geq \sum_{i=1}^{n} P(S_{n,i} < (1+c)Kn,\, Z_i < -Kn) - \frac{n(n-1)}{2} P(Z_i < -Kn)^2 \\
&= nP(Q_{n,i} < (1+c)Kn - (n-1)\nu) P(Z_i < -Kn) \\
&\quad - \frac{n(n-1)}{2} P(Z_i < -Kn)^2 \\
&= nP(Z_i < -Kn) \\
&\quad \cdot \left(P(Q_{n,i} < (1+c)Kn - (n-1)\nu) - \frac{n-1}{2} P(Z_i < -Kn) \right)
\end{aligned}$$

Now, as long as we choose $K > \frac{\nu}{c+1}$, we have that:

$$\log \left[P(Q_{n,i} < (1+c)Kn - (n-1)\nu) - \frac{n-1}{2} P(Z_i < -Kn) \right] \xrightarrow{P} 0$$

by the weak law of law of large numbers. Thus, in this case, we have that:

$$\lim_{n \to \infty} \frac{\log P(T_n < C_\alpha)}{R(n)} \geq \lim_{n \to \infty} \frac{\log nP(Z_i < -Kn)}{R(n)} = -1 \tag{21}$$

where we have used the fact that $\log P(Y_i < -Kn)$ is slowly-varying, Lemma 9, and X_i is subexponential.

(Case 2) In this case we have that $p \in (0, 1]$. Here, recall that we require $C_\alpha > 1$. Note that if $C_\alpha > 1$, then $c > 1$ as long as $n > 1 + c/\sqrt{c^2 - 1}$. Now, similar to before, we have

that, for any $K > 0$,

$$\begin{aligned}
P(T_n < C_\alpha) = P(S_n < cV_n) &\geq P(S_n < cV_n, \max_i Z_i > Kn) \\
&\geq P(S_n < c\max_i |Z_i|, \max_i Z_i > Kn) \\
&\geq P(S_n < c\max_i Z_i, \max_i Z_i > Kn) \\
&\geq \sum_{i=1}^{n} P(S_n < c\max_i |Z_i|, Z_i > Kn) \\
&\quad - \sum_{i \neq j} P(S_n < c\max_i |Z_i|, Z_i > Kn, Z_j > Kn) \\
&\geq \sum_{i=1}^{n} P(S_{n,i} < c\max_i |Z_i| - Z_i, Z_i > Kn) \\
&\quad - \frac{n(n-1)}{2} P(Z_i > Kn)^2 \\
&\geq \sum_{i=1}^{n} P(S_{n,i} < (c-1)Z_i, Z_i > Kn) \\
&\quad - \frac{n(n-1)}{2} P(Z_i > Kn)^2 \\
&\geq nP(Z_i > Kn)\left(P(Q_{n,i} < (c-1)Kn - (n-1)\nu) - \frac{n-1}{2}P(Z_i > Kn)\right)
\end{aligned}$$

Thus, just as before, if $K > \frac{\nu}{c-1}$,

$$\log\left[P(Q_{n,i} < (c-1)Kn - (n-1)\nu) - \frac{n-1}{2}P(Z_i > Kn)\right] \xrightarrow{P} 0$$

Therefore, (21) holds in this case as well, by a similar argument to that used in the previous case.

□

Next, we obtain an upper bound, summarized by the following lemma:

Lemma 2. *Assume the conditions of Theorem 1 hold. Then, we have that:*

$$\lim_{n\to\infty} \frac{\log P(T_n < C_\alpha)}{R(n)} \leq -1$$

To prove this lemma, we utilize the following decomposition:

$$
\begin{aligned}
P(T_n < C_\alpha) &= P\left(T_n < C_\alpha, \bigcap_i \{Z_i \in I_n\}\right) + P\left(T_n < C_\alpha, \bigcup_i \{Z_i \notin I_n\}\right) \\
&= U_n + P_n
\end{aligned}
$$

where I_n is some interval of the form $[-l_n, u_n]$, $l_n, u_n > 0$, $\frac{1}{n}l_n$, $\frac{1}{n}u_n \to q \in (0, \infty)$.

The following decomposition is quite useful. Let $I_n = [-l_n,\, u_n]$, $l_n, u_n > 0$, $l_n, u_n \to \infty$. Then, we have that:

$$
\begin{aligned}
P(T_n > C_\alpha) &= P\left(T_n > C_\alpha, \bigcap_i \{Z_i \in I_n\}\right) + P\left(T_n > C_\alpha, \bigcup_i \{Z_i \notin I_n\}\right) \\
&= U_n + P_n
\end{aligned}
$$

Bounding $\log P_n$ is straightforward. We will assume the same conditions used in Theorem 1, however note that in particular the requirement $C_\alpha > 1$ is not used here when constructing the upper bounds, and therefore it is an open question of whether that requirement is necessary in obtaining the type-II error convergence rates in this paper.

Lemma 3. *Under the conditions of Theorem 1, we have that:*

$$
\lim_{n\to\infty} \frac{\log P_n}{R(n)} \leq -1
$$

Proof. We use a simple upper found for P_n to start; for any $K > 0$, we have:

$$
\begin{aligned}
P_n &= P\left(T_n > C_\alpha, \bigcup_i \{Z_i \notin I_n\}\right) \\
&\leq P\left(\bigcup_i \{Z_i \notin I_n\}\right) \\
&\leq nP(Z_i \notin I_n) \\
&= nP(Z_i < -l_n) + nP(Z_i > u_n) \\
&= nP\left(Z_i < -n\frac{l_n}{n}\right) + nP\left(Z_i > n\frac{u_n}{n}\right)
\end{aligned}
$$

The result follows from Lemma 9 and the slow-variation of $R(n)$ and $\log P(|Z_i| > n)$. $\square$

Now, we need to bound U_n. As mentioned at the beginning of the appendix, the exponential inequalities used here draw heavily from such papers on large deviations for heavy tailed sums, such as Cline and Hsing (1989), Mikosch and Nagaev (1998), and Brosset et al. (2022). An interpretation of the following lemma is that truncated heavy-tailed random variables behave like thinner-tailed variables, even when the truncation point is diverging to infinity at the n-rate. In particular, in Mikosch and Nagaev (1998), in the proof of Theorem 6.1, it is clear that for random variables with regularly-varying tails, the truncated versions behave like random variables with higher tail indices, as far as tail-properties are concerned.

Lemma 4. *Under the conditions of Theorem 1, we have that:*

$$\lim_{n\to\infty} \frac{\log U_n}{R(n)} \leq -1$$

Proof. We first split up U_n further; for any $\omega \in (0, 1)$,

$$\begin{aligned}
U_n &= P\left(T_n > C_\alpha, \bigcap_i \{Z_i \in I_n\}\right) \\
&= P\left(S_n < cV_n, \bigcap_i \{Z_i \in I_n\}\right) \\
&= P\left(-Q_n > \omega n\nu, \bigcap_i \{Z_i \in I_n\}\right) \\
&\qquad + P\left(V_n > \frac{1-\omega}{c} n\nu, \bigcap_i \{Z_i \in I_n\}\right) \\
&= U_{n1} + U_{n2}
\end{aligned}$$

We will deal with U_{n2}, as U_{n1} can be handled in a similar fashion. Let $U_i = Z_i^2$. Then, we will let $l_n = u_n$, so that $Z_i \in I_n$ if and only if $U_i \leq u_n^2$. Let $y_n = \frac{1-\omega}{c} n\nu$, and G be the distribution function of U_i, and notice that since $\log(1 - F(u) + F(-u))$ is slowly varying, so is $\log(1-G(u^2))$. Using Markov's inequality, for some sequence $s_n \geq 0$, with $s_n \to 0$,

$$
\begin{aligned}
U_{n2} &\leq e^{-s_n y_n^2} \left(\int_0^{u_n^2} e^{s_n u} G(du) \right)^n \\
&= e^{-s_n y_n^2} \left(\int_0^{\beta/\sqrt{s_n}} e^{s_n u} G(du) + \int_{\beta/\sqrt{s_n}}^{u_n^2} e^{s_n u} G(du) \right)^n \\
&= e^{-s_n y_n^2} (I_{21} + I_{22})^n
\end{aligned}
$$

for $\beta \in (0,1)$. Consider first I_{21}. For $u \in (0, \beta/\sqrt{s_n})$, we have that $e^{s_n u} \in (1, e^{\beta\sqrt{s_n}})$. The convex function $e^{s_n u}$ is bounded above by the secant:

$$
1 + \frac{e^{\beta\sqrt{s_n}} - 1}{\beta} \sqrt{s_n} u = 1 + K_n \sqrt{s_n} u
$$

Note that $K_n = O(1/\sqrt{s_n})$. Thus, Thus, we have that:

$$
I_{12} \leq \mathbb{E}(1 + K_n \sqrt{s_n} U) = 1 + K_n \sqrt{s_n}(1 + \nu^2) = 1 + o(s_n)
$$

Now, for I_{22}, we use integration by parts:

$$
\begin{aligned}
I_{22} = &- \left(e^{s_n u}(1 - G(u))\right|_{\beta/\sqrt{s_n}}^{u_n^2} + s_n \int_{\beta/\sqrt{s_n}}^{u_n^2} e^{s_n u}(1 - G(u)) du \\
&\leq e^{\beta}(1 - G(\beta/\sqrt{s_n})) + s_n \int_{\beta/\sqrt{s_n}}^{u_n^2} e^{s_n u + \log(1 - G(u))} du \\
&\leq e^{\beta}(1 - G(\beta/\sqrt{s_n})) + s_n \int_{\beta/\sqrt{s_n}}^{u_n^2} e^{s_n u + \log(1 - G(u))} du
\end{aligned}
$$

Since $\log(1 - G(u))$ is a slowly varying, monotonically decreasing function and $s_n > 0$, $s_n u + \log(1 - G(u))$ is convex on $(\beta/\sqrt{s_n},\, u_n^2)$ for n large enough, and therefore achieves its maximum at a boundary point. We would like to show that:

$$
s_n u_n^2 + \log(1 - G(u_n^2)) < \beta\sqrt{s_n} + \log(1 - G(\beta/\sqrt{s_n})) \tag{22}
$$

holds for n large enough. If we set $s_n = \frac{R(n)}{u_n^2}$, since $(1 - G(u_n^2))/(1 - G(\beta u_n/\sqrt{R(n)})) \to 0$, we have that for large enough n,

$$\begin{aligned}
I_{22} &\leq e^{\beta}(1 - G(\beta/\sqrt{s_n})) + s_n \left(u_n^2 - \frac{\beta}{\sqrt{s_n}}\right) e^{\beta\sqrt{s_n} + \log(1 - G(\beta/\sqrt{s_n}))} \\
&= (1 - G(\beta/\sqrt{s_n})) \left(e^{\beta} + (s_n u_n^2 - \beta\sqrt{s_n}) e^{\beta\sqrt{s_n}}\right) \\
&= o(\sqrt{R(n)}/n)
\end{aligned}$$

This all also implies that $I_{12} = 1 + o(1/n)$. Putting this all together, if we choose $u_n = w y_n$, for $w \in (0, 1)$,

$$\begin{aligned}
\lim_{n\to\infty} \frac{\log U_n}{R(n)} &\leq \lim_{n\to\infty} \left\{ -\frac{y_n^2}{u_n^2} + \frac{1}{\sqrt{R(n)}} \frac{n}{\sqrt{R(n)}} \log(1 + O(\sqrt{R(n)}/n)) \right\} \\
&\to -\frac{1}{w^2} < -1
\end{aligned}$$

The proof for U_{n1} is similar. ∎

The overall upper bound is achieved by combining Lemma 3 and 4:

Lemma 5. *Under the conditions of Theorem 1, we have that:*

$$\limsup_{n\to\infty} \frac{\log P(T_n < C_\alpha)}{R(n)} = -1$$

Proof. Use Lemmas 3 and 4 along with Lemma 8. ∎

A.2 Proof of Theorem 2

We will apply Lemmas 11 and 12, as appropriate, to U-statistics of the form:

$$U_{n,m} = \frac{1}{\binom{n}{m}} \sum_{j \in \mathcal{A}_{n,m}} f(T_{j,m} - \sqrt{m/n}\delta) - \mathbb{E}\, f(T_{j,m} - \sqrt{m/n}\delta)$$

where $\mathbb{E}\, X_i = \delta/\sqrt{n}$. Thus, $h_m(X_{j_1}, \ldots, X_{j_m}) = f(T_{j,m} - \sqrt{m/n}\delta) - \mathbb{E}\, f(T_{j,m} - \sqrt{m/n}\delta)$. In our case, we will have that $f(T_{j,m} - \sqrt{m/n}\delta) \Rightarrow f(Z)$, $Z \sim \mathcal{N}(0, 1)$. We will need to establish the asymptotic behavior of $\sigma_{1,m}^2$ and $\sigma_{m,m}^2$ to apply these lemmas in our cases of interest. We abbreviate when there is no loss in clarity $f_m(T_{j,m}) := f(T_{j,m} - \sqrt{m/n}\delta)$. For the next lemma, we define:

$$T_m(X_{-j_1}, z) = \frac{\sqrt{m}\bar{X}_{-j_1} + \frac{1}{\sqrt{m}}(z - \bar{X}_{-j_1})}{\sqrt{\frac{m-2}{m-1}S_{-j_1}^2 + \frac{1}{m}(z - \bar{X}_{-j_1})^2}}$$
$$= \sqrt{\frac{m}{m-1}}\frac{\sqrt{m-1}\bar{X}_{-j_1}}{S_{-j_1}}\left(\frac{1}{\sqrt{\frac{m-2}{m-1} + \frac{1}{m}\left(\frac{z-\bar{X}_{-j_1}}{S_{-j_1}}\right)^2}}\right) + \frac{\frac{1}{\sqrt{m}}\frac{z-\bar{X}_{-j_1}}{S_{-j_1}}}{\sqrt{\frac{m-2}{m-1} + \frac{1}{m}\left(\frac{z-\bar{X}_{-j_1}}{S_{-j_1}}\right)^2}}$$

Clearly, $T(X_{-j_1}, X_{j_1}) = T_{j,m}$. Now, we also define:

$$g_{f_m}(z) := (f_m \circ T)(X_{-j_1}, \sqrt{m}z + \bar{X}_{-j_1})$$
$$h_m := \frac{X_{j_1} - \bar{X}_{-j_1}}{\sqrt{m}}$$
$$\tilde{h}_m := \frac{\tilde{X} - \bar{X}_{-j_1}}{\sqrt{m}}$$

where $\tilde{X}$ is independent of and identically distributed as the other X_i.

Lemma 6. *Assume that for some $d \in (0,1)$, $\sup_m \mathbb{E}\,|f_m(T_{j,m})|^{1+d} = D < \infty$, and $\|f'(z)\|_\infty < \infty$. Then, there exists a function $\eta(z)$ such that for each X_{j_1},*

$$\sup_{m>2} \mathbb{E}\left(\sqrt{m}|g_{f_m}(h_m) - g_{f_m}(\tilde{h}_m)|^{1+d'}|X_{j_1}\right) \leq \eta(X_{j_1}) < \infty, \text{ a.s.}$$

Proof. We drop the explicit dependence of g on f_m for notational simplicity. Without loss of generality let $\mathrm{Var}(X_i) = 1$. Let c be such that $P(S_{-j_1}^2 < c) \leq \exp\{-(n-1)K_c\}$. Existence of such a c comes from Lemma 10. We also define $A_{m,\epsilon} = [|h_m| < 1/M_\epsilon,\ |\tilde{h}_m < 1/M_\epsilon]$, where M_ϵ is defined such that if $\max\{h_m, \tilde{h}_m\} < 1/M_\epsilon$, then

$$\max_{z \in \{h_m, \tilde{h}_m\}} \left\{\left|\frac{g(z) - g(0)}{z} - g'(0)\right|\right\} \leq \epsilon.$$

Let $d' \in (0, d)$. First, we write:

$$\begin{aligned}
|\sqrt{m}(g(h_m) - g(\tilde{h}_m))|^{1+d'} &= |\sqrt{m}(g(h_m) - g(\tilde{h}_m))|^{1+d'} \mathbb{1}_{[S^2_{-j_1} < c]} \\
&\quad + |\sqrt{m}(g(h_m) - g(\tilde{h}_m))|^{1+d'} \mathbb{1}_{[S^2_{-j_1} \geq c]} \mathbb{1}_{A_{m,\epsilon}} \\
&\quad + |\sqrt{m}(g(h_m) - g(\tilde{h}_m))|^{1+d'} \mathbb{1}_{[S^2_{-j_1} \geq c]} \mathbb{1}_{A^c_{m,\epsilon}} \\
&= G_1 + G_2 + G_3
\end{aligned}$$

First, we consider $\mathbb{E}(G_1|X_{j_1})$. Using Hölder's inequality and Lemma 10, we have

$$\begin{aligned}
\mathbb{E}(G_1|X_{j_1}) &\leq \left(\mathbb{E}(|g(h_m) - g(\tilde{h}_m)|^{1+d}|X_{j_1})\right)^{\frac{1+d'}{1+d}} m^{1/2+d'/2} P(S^2_{-j_1} < c) \\
&\leq \mathbb{E}(|g(h_m) - g(\tilde{h}_m)|^{1+d}|X_{j_1})^{\frac{1+d'}{1+d}} \frac{1+d'}{2K_c} \\
&\leq \mathbb{E}\left(|g(h_m)|^{1+d}|X_{j_1}\right)^{\frac{1+d'}{1+d}} D \frac{1+d'}{2K_c}
\end{aligned}$$

Next, we consider $\mathbb{E}(G_2|X_{j_1})$. For this event, we rewrite:

$$g(h_m) - g(\tilde{h}_m) = g(h_m) - g(0) - g'(0)h_m - (g(\tilde{h}_m) - g(0) - g'(0)\tilde{h}_m) + g'(0)(h_m - \tilde{h}_m)$$

On $A_{m,\epsilon}$, we have that:

$$\begin{aligned}
|g(h_m) - g(0) - g'(0)h_m| &\leq \epsilon h_m \\
|g(\tilde{h}_m) - g(0) - g'(0)\tilde{h}_m| &\leq \epsilon \tilde{h}_m
\end{aligned}$$

Furthermore, we have as an expression for $g'(0)$

$$g'(0) = \frac{1}{\sqrt{\frac{m-2}{m-1} S^2_{-j_1}}} f'\left(\sqrt{\frac{m}{m-2}} \frac{\sqrt{m-1}\bar{X}_{-j_1}}{\sqrt{S^2_{-j_1}}}\right) \leq \frac{1}{\sqrt{\frac{m-2}{m-1} c}} \|f'(z)\|_\infty$$

which implies that:

$$|g'(0)(h_m - \tilde{h}_m)| \leq \frac{1}{\sqrt{\frac{m-2}{m-1} c}} \|f'(z)\|_\infty |h_m - \tilde{h}_m|$$

Combining these bounds, we have:

$$
\begin{aligned}
\mathbb{E}(G_2|X_{j_1}) &\leq \epsilon^{1+d'}\,\mathbb{E}(|\sqrt{m}h_m|^{1+\delta'}|X_{j_1}) \\
&\quad + \epsilon^{1+d'}\,\mathbb{E}(|\sqrt{m}\tilde{h}_m|^{1+\delta'}) \\
&\quad + \frac{\sqrt{m-1}}{\sqrt{(m-2)c}}\|f'(z)\|_\infty\,\mathbb{E}(|\sqrt{m}(h_m-\tilde{h}_m)|^{1+d'}|X_{j_1}) \\
&\leq \epsilon^{1+d'}\left(\mathbb{E}(|X_{j_1}-\bar{X}_{j-1}|^{1+d'}|X_{j_1}) + \mathbb{E}\,|\tilde{X}-\bar{X}_{j-1}|^{1+d'}\right) \\
&\quad + \frac{\sqrt{2}}{\sqrt{c}}\|f'(z)\|_\infty\,\mathbb{E}\,|X_{j_1}-\tilde{X}|^{1+d'}
\end{aligned}
$$

Next, we consider $\mathbb{E}(G_3|X_{j_1})$. In this case, we have:

$$
\mathbb{E}(G_3|X_{j_1}) \leq \mathbb{E}(|(g(h_m)-g(\tilde{h}_m))|^{1+d}|X_{j_1})m^{1/2+d'/2}P(A^c_{m,\epsilon})
$$

We now examine $P(A^c_{m,\epsilon})$ and apply Chebyshev's inquality:

$$
\begin{aligned}
P(A^c_{m,\epsilon}|X_{j_1}) &\leq P(|X_{j_1}-\bar{X}_{-j_1}| > \sqrt{m}/M_\epsilon|X_{j_1}) + P(|X-\bar{X}_{-j_1}| > \sqrt{m}/M_\epsilon) \\
&\leq \frac{M_\epsilon^2}{m^{1/2+d'/2}}\left(\mathbb{E}(|\bar{X}_{-j_1}-X_{j_1}|^{1+d'}|X_{j_1}) + \mathbb{E}(|\bar{X}_{-j_1}-X|^{1+d'}\right)
\end{aligned}
$$

Thus,

$$
\mathbb{E}(G_3|X_{j_1}) \leq \mathbb{E}(|g(h_m)-g(\tilde{h}_m)|^{1+d}|X_{j_1})M_\epsilon^2\left(\mathbb{E}(|\bar{X}_{-j_1}-X_{j_1}|^{1+d'}|X_{j_1}) + \mathbb{E}(|\bar{X}_{-j_1}-\tilde{X}|^{1+d'}\right)
$$

Combining the bounds on $\mathbb{E}(G_1|X_{j_1})$, $\mathbb{E}(G_2|X_{j_1})$, and $\mathbb{E}(G_3|X_{j_1})$ completes the proof. □

Lemma 7. *Assume that for some $d > 0$, there is some D such that $\sup_m |f_m(T_{j,m})|^{2+d} \leq D < \infty$, $\mathbb{E}\,X_i = \delta/\sqrt{n}$, $T_{j,m} - \sqrt{m/n}\delta \Rightarrow Z \sim \mathcal{N}(0,1)$, and $\|f'(z)\|_\infty < \infty$. Then,*

$$
\sigma^2_{m,m} = \mathrm{Var}(f_m(T_{j,m})) \to \mathrm{Var}(f(Z)) \tag{23}
$$

$$
m\sigma^2_{1,m} = m\,\mathrm{Var}(\mathbb{E}(f_m(T_{j,m})|X_{j_1})) \to (\mathbb{E}\,f'(Z))^2 \tag{24}
$$

Proof. The techniques here a quite similar to those in Lemma 6. Without loss of generality let $\mathrm{Var}(X_i) = 1$. The first part, (23) is immediate since the $f_m(T_{j,m})^2$ are uniformly integrable by assumption. We will abbreviate $dP^{\otimes}(X_{-j_1}) = dP(X_{j_2}) \times \cdots dP(X_{j_m})$. To get (24), consider that:

$$
\begin{aligned}
m\,\mathrm{Var}(\mathbb{E}(f_m(T_{j,m})|X_{j_1})) &= \mathbb{E}[(\sqrt{m}\,\mathbb{E}(f_m(T_{j,m})|X_{j_1}) - \mathbb{E}\,f_m(T_{j,m}))^2] \\
&= \int \left(\sqrt{m}\int g(h_m) - g(\tilde{h}_m) dP^{\otimes}(X_{-j_1})\right)^2 dP(X_{j_1}) \qquad (25)
\end{aligned}
$$

We would like to apply Vitali's Convergence Theorem. Therefore, we will aim to bound

$$
\left(\sqrt{m}\int g(h_m) - g(\tilde{h}_m) dP^{\otimes}(X_{-j_1})\right)^2 \qquad (26)
$$

by an integrable function. First, consider the event $B_m := [S_{-j_1} < c]$. By Lemma 10, we can choose a c such that $P(B_m) \leq \exp\{-mK_c\}$, for some $K > 0$, which does not depend on m. Consider then, if we split up the integral in (26) over B_m, that by Hölder's inequality,

$$
\begin{aligned}
\left(\sqrt{m}\int_{B_m} g(h_m) - g(\tilde{h}_m) dP^{\otimes}(X_{-j_1})\right)^2 &\leq \left(\sqrt{m}\int |g(h_m) - g(\tilde{h}_m)|^{\frac{2+d}{2}} dP^{\otimes}(X_{-j_1}) P(B_m)\right)^2 \\
&\leq \int |g(h_m) - g(\tilde{h}_m)|^{2+d} dP^{\otimes}(X_{-j_1}) \cdot mP(B_m)^2 \\
&\leq \int |g(h_m) - g(\tilde{h}_m)|^{2+d} dP^{\otimes}(X_{-j_1}) \cdot \frac{1}{2K_c}
\end{aligned}
$$

This is integrable since $\sup_m \mathbb{E}\,f_m(T_{j,m})^{2+d} \leq D < \infty$. Thus, we can focus our attention on B_m^c, i.e. the case in which $S_{-j_1}^2 \geq c$. We omit including B_m^c explicitly in our regions of integration in what follows for notational convenience. On this event, we rewrite

$$
g(h_m) - g(\tilde{h}_m) = g(h_m) - g(0) - g'(0)h_m - (g(\tilde{h}_m) - g(0) - g'(0)\tilde{h}_m) + g'(0)(h_m - \tilde{h}_m).
$$

Let $\epsilon > 0$. We can then choose M_ϵ such that on the event $A_{m,\epsilon} = [|h_m| \leq 1/M_\epsilon, |\tilde{h}_m| \leq 1/M_\epsilon]$, we have that:

$$
\begin{aligned}
\sqrt{m}|g(h_m) - g(0) - g'(0)h_m| &\leq \epsilon\sqrt{m}|h_m| \\
\sqrt{m}|g(\tilde{h}_m) - g(0) - g'(0)\tilde{h}_m| &\leq \epsilon\sqrt{m}|\tilde{h}_m|
\end{aligned}
$$

Now, note that:

$$\left(\sqrt{m}\int_{A_{m,\epsilon}} g(h_m) - g(\tilde{h}_m) dP^{\otimes}(X_{-j_1})\right)^2 \tag{27}$$
$$\leq \left(\sqrt{m}\int_{A_{m,\epsilon}} |g(h_m) - g(\tilde{h}_m)| dP^{\otimes}(X_{-j_1})\right)^2$$
$$\leq \left(\int_{A_{m,\epsilon}} \epsilon\sqrt{m}(|h_m| + |\tilde{h}_m|) + |g'(0)||h_m - \tilde{h}_m| dP^{\otimes}(X_{-j_1})\right)^2$$

Thus, if we can bound each of

$$\int_{A_{m,\epsilon}} m h_m^2 dP^{\otimes}(X_{-j_1}),\ \int_{A_{m,\epsilon}} m\tilde{h}_m^2 dP^{\otimes}(X_{-j_1}),\ \int_{A_{m,\epsilon}} m g'(0)^2 (h_m - \tilde{h}_m)^2 dP^{\otimes}(X_{-j_1})$$

then we have bounded (27). The first two integrals are bounded since the observations X_i have finite variance:

$$\mathbb{E}(m h_m^2 | X_{j_1}) = \mathbb{E}(X_{j_1} - \bar{X}_{-j_1})^2 = (X_{j_1} - \delta/\sqrt{n})^2 + \frac{1}{m},\ \mathbb{E}\, m\tilde{h}_m^2 = 1 + \frac{1}{m}$$

For the third term, note that $m(h_m - \tilde{h}_m)^2 = (X_{j_1} - \tilde{X})^2$. For $g'(0)$, since we are considering only the case when $S_{-j_1}^2 > c$, we have that

$$g'(0) = \frac{1}{\sqrt{\frac{m-2}{m-1} S_{-j_1}^2}} f'\left(\sqrt{\frac{m}{m-2}}\frac{\sqrt{m-1}\bar{X}_{-j_1}}{\sqrt{S_{-j_1}^2}}\right) \leq \frac{1}{\sqrt{\frac{m-2}{m-1} c}} \|f'(z)\|_\infty$$

Therefore, we can apply Cauchy Schwarz and say:

$$\int_{A_{m,\epsilon}} g'(0)^2 (h_m - \tilde{h}_m)^2 dP^{\otimes}(X_{-j_1}) \leq \frac{m-1}{m-2}\frac{1}{c}\|f'(z)\|_\infty^2\, \mathbb{E}((h_m - \tilde{h}_m)^2 | X_{j_1})$$
$$\leq \frac{m-1}{m-2}\frac{1}{c}\|f'(z)\|_\infty^2 (X_{j_1} - \tilde{X})^2$$

Lastly, we consider $A_{m,\epsilon}^c = [h_m > K_\epsilon] \cup [\tilde{h}_m > K_\epsilon]$. In this case we have:

$$\left(\sqrt{m}\int_{A_{m,\epsilon}^c} g(h_m) - g(\tilde{h}_m) dP^{\otimes}(X_{-j_1})\right)^2$$
$$\leq m P(A_{m,\epsilon}^c)^2 \int |g(h_m) - g(\tilde{h}_m)|^{2+d} dP^{\otimes}(X_{-j_1})$$

Thus, as long as $P(A^c_{m,\epsilon}) = O(1/\sqrt{m})$ we are done. To this effect, consider first that $P(A^c_{m,\epsilon}) \le P(|h_m| > 1/M_\epsilon) + P(|\tilde{h}_m > 1/M_\epsilon)$. Considering the first probability in this sum, we have by Chebyshev's inequality:

$$\begin{aligned}
P(|h_m| > 1/M_\epsilon) &\le P(|X_{j_1} - \delta/\sqrt{n}| > \sqrt{m}/2M_\epsilon) + P(|\bar{X}_{-j_1} - \delta/\sqrt{n}| > \sqrt{m}/2M_\epsilon) \\
&\le \frac{4M_\epsilon^2}{m} + \frac{4M_\epsilon^2}{m^2}
\end{aligned}$$

Similarly, for $\tilde{h}_m$, thus we have that:

$$\left(\sqrt{m}\int g(h_m) - g(\tilde{h}_m) dP^{\otimes}(X_{-j_1})\right)^2 \le \bar{\eta}(X_{j_1})$$

where $\mathbb{E}\,|\bar{\eta}(X_{j_1})| < \infty$. Thus, we can apply Vitali's Convergence Theorem:

$$\lim_{m\to\infty} m \operatorname{Var}(\mathbb{E}(f(T_{j,m})|X_{j_1})) = \mathbb{E}(\lim_{m\to\infty}(\sqrt{m}\,\mathbb{E}(f(T_{j,m})|X_{j_1}) - \mathbb{E}\, f(T_{j,m}))^2)$$

Now, we apply Lemma 6, and have:

$$\mathbb{E}(\lim_{m\to\infty}(\sqrt{m}\,\mathbb{E}(f(T_{j,m})|X_{j_1}) - \mathbb{E}\, f(T_{j,m}))^2) = \mathbb{E}(\mathbb{E}(\lim_{m\to\infty}(\sqrt{m}(f(T_{j,m}) - \mathbb{E}\, f(T_{j,m})))|X_{j_1}))^2$$

Thus, note that, conditional on X_{j_1}, by the definition of the derivative,

$$\sqrt{m}(f(T_{j,m}) - \mathbb{E}\, f(T_{j,m})) \Rightarrow f'(Z)(X_{j_1} - \delta/\sqrt{n})$$

where Z and X_{j_1} are independent. Thus, $m\sigma^2_{1,m} \to (\mathbb{E}\, f'(Z))^2$, as desired. □

We are now ready to prove Theorem **??**

Proof. We build a locally-quadratic approximation to the convex objective function. Let:

$$\begin{aligned}
A_{n,m}(t) &= \frac{1}{\binom{n}{m}} \sum_j \rho\left(T_{j,m} - \sqrt{\frac{m}{n}}(t+\delta)\right) \\
U_{n,m}(t) &= \frac{1}{\binom{n}{m}} \sum_j \frac{1}{\sqrt{m}} \rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}(t+\delta)\right) \\
J_{n,m}(t) &= \frac{1}{\binom{n}{m}} \sum_j \rho''\left(T_{j,m} - \sqrt{\frac{m}{n}}(t+\delta)\right)
\end{aligned}$$

Note that $A_{n,m}(t)$ is minimized at $\sqrt{n}\hat{\tau} - \delta$. Now, for each fixed t, we have:

$$A_{n,m}(t) - A_{n,m}(0) = -\frac{m}{\sqrt{n}} U_{n,m}(0)t + \frac{1}{2}\frac{m}{n} J_{n,m}(0)t^2 + O\left((m/n)^{3/2}\right)$$

We will use the approach in Hjort and Pollard (2011). Note that $(m/n)(A_{n,m}(t) - A_{n,m}(0))$ is a convex function in t which is minimized at $\sqrt{n}\hat{\tau} - \delta$. Now, consider that since

$$(n/m)(A_{n,m}(t) - A_{n,m}(0)) + \sqrt{n}U_{n,m}(0)t - \frac{1}{2}J_{n,m}(0)t^2 = O\left((m/n)^{1/2}\right)$$

by the Basic Corollary of Hjort and Pollard (2011), the minimizer of $(n/m)(A_{n,m}(t) - A_{n,m}(0))$, $\sqrt{n}\hat{\tau} - \delta$, can be represented as:

$$\begin{aligned}\sqrt{n}\hat{\tau} - \delta &= \arg\min_t \left\{\sqrt{n}U_{n,m}(0)t - \frac{1}{2}J_{n,m}(0)t^2\right\} + o_P(1) \\ &= \frac{\sqrt{n}U_{n,m}(0)}{J_{n,m}(0)} + o_P(1) \end{aligned} \tag{28}$$

First, note that since $\rho''(z)$ is bounded, $J_{n,m}(0)$ is uniformly integrable and converges in mean square to $\mathbb{E}\,\rho''(Z)$. For the numerator, consider the U-statistic:

$$U_{n,m}(0) - \mathbb{E}\,U_{n,m}(0) = \frac{1}{\binom{n}{m}} \sum_{j \in \mathcal{A}_{n,m}} \frac{1}{\sqrt{m}} \left(\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right) - \mathbb{E}\,\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)\right)$$

We consider the following ratio, applying Lemma 7 (with $f = \rho'$):

$$\frac{\operatorname{Var}\left(\frac{1}{\sqrt{m}}\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)\right)}{m\operatorname{Var}\left(\mathbb{E}\left(\frac{1}{\sqrt{m}}\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)|X_{j_1}\right)\right)} = \frac{\operatorname{Var}\left(\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)\right)}{m\operatorname{Var}\left(\mathbb{E}\left(\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)|X_{j_1}\right)\right)} \to \frac{\operatorname{Var}(\rho'(Z))}{(\mathbb{E}\,\rho''(Z))^2} \tag{29}$$

Now, note that, also by Lemma 7,

$$\begin{aligned} m\sqrt{\operatorname{Var}\left(\frac{1}{\sqrt{m}}\mathbb{E}\left(\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)|X_{j_1}\right)\right)} &= \sqrt{m\operatorname{Var}\left(\mathbb{E}\left(\rho'\left(T_{j,m} - \sqrt{\frac{m}{n}}\delta\right)|X_{j_1}\right)\right)} \\ &\to \mathbb{E}\,\rho''(Z) \end{aligned}$$

We then have, by Lemma 11 and Slutsky's Theorem, that:

$$\frac{\sqrt{n}(U_{n,m}(0) - \mathbb{E}\,U_{n,m}(0))}{J_{n,m}(0)} \Rightarrow \mathcal{N}(0,1) \tag{30}$$

Thus, we re-center $\hat{\tau}$ in (28):

$$\sqrt{n}\left(\hat{\tau} - \frac{\mathbb{E}\,\rho'(T_{j,m})}{\sqrt{m}\,\mathbb{E}\,\rho''(T_{j,m})}\right)$$
$$= \delta + \frac{\sqrt{n}(U_{n,m}(0) - \mathbb{E}\,U_{n,m}(0))}{J_{n,m}(0)} + \sqrt{\frac{n}{m}}\,\mathbb{E}\,\rho'(T_{j,m})\left(\frac{1}{J_{n,m}(0)} - \frac{1}{\mathbb{E}\,\rho''(T_{j,m})}\right)$$

To show that the last term is negligible, first recall that since $\mathrm{Var}(\rho''(T_{j,m})) < \infty$, we have that

$$\mathrm{Var}(J_{n,m}(0)) \leq \frac{m}{n}\,\mathrm{Var}(\rho''(T_{j,m}))$$

This implies that $J_{n,m}(0) - \mathbb{E}\,\rho''(T_{j,m}) = O_P\left(\sqrt{\frac{m}{n}}\right)$, and as $m/n \to 0$, this completes the proof.

□

A.3 Proof of Theorem 3

Here we provide a simple proof of the exponential bound for Type-II error probabilities using the new test.

Proof. Let:

$$\Psi_{n,m}(t) = \frac{1}{\binom{n}{m}} \sum_{j \in \mathcal{A}_{n,m}} \rho'\left(T_{j,m} - \sqrt{m}t\right)$$

Consider that the probability of a Type-II error is:

$$\begin{aligned}
P(\sqrt{n}\hat{\tau} < C_\alpha) &= P\left(0 < -\Psi_{n,m}\left(C_\alpha/\sqrt{n}\right)\right) \\
&= P\left(1/2 < -\Psi_{n,m}\left(C_\alpha/\sqrt{n}\right)/2\|\rho'\|_\infty + 1/2\right) \\
&\leq P\left(1/2 < \frac{1}{\binom{n}{m}} \sum_{j \in \mathcal{A}_{n,m}} \mathbb{1}_{[T_{j,m} > \sqrt{m/n}C_\alpha]}\right)
\end{aligned}$$

Let $p_m = P(T_{j,m} > \sqrt{m/n}C_\alpha)$. By Hoeffding's inequality for U-statistics, we have:

$$\begin{aligned}
-\frac{m}{n}\log P(\sqrt{n}\hat{\tau} < C_\alpha) &\geq -\frac{m}{n}\log 2 + (1/2 - p_m)^2\frac{n}{m} \\
&= -\frac{m}{n}\log 2 + \frac{1}{4} + p_m^2 - p_m \\
&\to \frac{1}{4}
\end{aligned}$$

□

B Technical Appendix

This first pair of lemmas is useful in proving Theorem 10. First, we make use of the following well known result:

Lemma 8 (Dembo and Zeitouni (2009), Lemma 1.2.15)**.** *Let a_n be a sequence of positive constants such that $\lim_{n\to\infty} a_n = \infty$. Let N be a fixed integer, and let $\Pi_{i,n}$, $i = 1, \ldots, N$ be a sequence of non-negative constants. Then, we have that:*

$$\limsup_{n\to\infty} \frac{1}{a_n} \log\left(\sum_{i=1}^{n} \Pi_{i,n}\right) = \max_{1\le i\le N} \limsup_{n\to\infty} \frac{1}{a_n} \log \Pi_{i,n} \tag{31}$$

The next lemma is closely related to Lemma 8.

Lemma 9. *Let A_n, B_n be two sequences non-negative sequences such that $\log B_n \to \pm\infty$ and $A_n/B_n \to q \in [0,\infty)$. Then, we have that:*

$$\lim_{n\to\infty} \frac{\log(A_n + B_n)}{\log B_n} = 1 \tag{32}$$

Proof. By rearranging, we have:

$$\begin{aligned}
\frac{\log(A_n + B_n)}{\log B_n} &= \frac{\log\left(1 + \frac{A_n}{B_n}\right) + \log B_n}{\log B_n} \\
&= 1 + \frac{\log\left(1 + \frac{A_n}{B_n}\right)}{\log B_n} \\
&\to 1
\end{aligned}$$

□

The previous lemma is useful for us when A_n and B_n are sequences of tail probabilities. When B_n corresponds to tail probabilities from the heavier of the two tails, this lemma says that to characterize the limiting behavior of both tail probabilities we only need consider the contribution of the heavier tail. Notice that we use "heavier" in a weak sense: we allow $A_n/B_n \to c \in (0,\infty)$, in which case A_n and B_n have comparable limiting behavior.

For proving Theorem 2, we will also use several technical lemmas. This lemma is likely well-known but we could not find a proof and therefore provide one here:

Lemma 10. *Assume that there exists a* $c > 0$ *such that* $0 < c < \mathbb{E}\, X_i^2$*. Then, there exists a* $\epsilon > 0$ *such that:*

$$P(S_n^2 < \epsilon) \leq \exp\{-nK_\epsilon\}$$

for some $K_\epsilon > 0$.

Proof. Since S_n^2 is shift-invariant, without loss of generality assume $\mathbb{E}\, X_i = 0$. Note the simple fact that:

$$\begin{aligned}\frac{(X_i - X_j)^2}{2} - \delta &:= Y_{ij,\delta} \\ &:= Y_{ij,\delta}^+ - Y_{ij,\delta}^- \\ &\geq -Y_{ij,\delta}^-\end{aligned}$$

Furthermore, consider that:

$$\begin{aligned}\mathbb{E}\, Y_{ij,\delta}^- &= \mathbb{E}\left(\frac{(X_i - X_j)^2}{2} \mathbb{1}_{[(X_i - X_j)^2 \leq 2\delta]}\right) \\ &\leq \delta P((X_i - X_j)^2 \leq 2\delta)\end{aligned}$$

Consider: if $P((X_i - X_j)^2 \leq 2\delta) = 1$, then we can apply the Hoeffding bound for U-statistics directly: for $\epsilon < \sigma^2$, we have:

$$P(S_n^2 < \epsilon) = P(-(S_n^2 - \sigma^2) > \sigma^2 - \epsilon) \leq \exp\left\{-\frac{n(\sigma^2 - \epsilon)^2}{\delta^2}\right\}$$

Now, consider the case that $P((X_i - X_j)^2 \leq 2\delta) < 1$. In this case, we can choose $0 < \epsilon < \delta - \mathbb{E}\, Y_{ij,\delta}^-$

$$\delta - \epsilon - \mathbb{E}\, Y_{ij,\delta}^- > 0$$

In this case, we have:

$$
\begin{aligned}
P(S_n^2 < \epsilon) &= P(S_n^2 - \delta < \epsilon - \delta) \\
&= P\left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} Y_{ij,\delta}^{+} - Y_{ij,\delta}^{-} < \epsilon - \delta\right) \\
&\leq P\left(-\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} Y_{ij,\delta}^{-} < \epsilon - \delta\right) \\
&= P\left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} Y_{ij,\delta}^{-} > \delta - \epsilon\right) \\
&= P\left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} (Y_{ij,\delta}^{-} - \mathbb{E}\, Y_{ij,\delta}^{-}) > \delta - \epsilon - \mathbb{E}\, Y_{ij,\delta}^{-}\right) \\
&\leq \exp\left\{-\frac{n(\delta - \epsilon - \mathbb{E}\, Y_{ij,\delta})^2}{\delta^2}\right\}
\end{aligned}
$$

□

Effectively the lemma says that deviations of the sample variance below the sample variance disappear exponentially quickly in n, and the reason for this is those are essentially coming from the contributions of bounded random variables.

The next two lemmas can be found int Peng et al. (2019) and the first is found in Minsker (2022); the second is mentioned but not proved in Peng et al. (2019). We include both for completeness.

Throughout this section, let $h_m(X_1, \ldots, X_m)$ be a symmetric function of m arguments, with $m \to \infty$ and $\mathbb{E}\, h_m(X_1, \ldots, X_m) = 0$. The U-statistic with kernel h_m is then:

$$
U_{n,m} := \frac{1}{\binom{n}{m}} \sum_{j \in \mathcal{A}_{n,m}} h_m(X_{j_1}, \ldots, X_{j_m})
$$

where as before we let $\mathcal{A}_{n,m} = \{j : \{j_1, \ldots, j_m\} \subset \{1, \ldots, n\}, j_k \neq j_l,\ l \neq k\}$, i.e. the set of all subsamples of size m of the set of indices $\{1, \ldots, n\}$. We define the following objects:

$$\begin{aligned}
\psi_{c,m}(X_1, \ldots, X_c) &:= \mathbb{E}[h_m(X_1, \ldots, X_m)|X_1, \ldots, X_c] \\
h_m^{(1)}(X_1) &:= \psi_{1,m}(X_1) \\
h_m^{(c)}(X_1, \ldots, X_c) &:= \psi_{c,m}(X_1, \ldots, X_c) - \sum_{k=1}^{c-1} \sum_{j \in \mathcal{A}_{c,k}} h_m^{(k)}(X_{j_1}, \ldots, X_{j_k}) \\
\delta_{c,m}^2 &:= \mathrm{Var}(h_m^{(c)}(X_1, \ldots, X_c)) \\
\sigma_{c,m}^2 &:= \mathrm{Var}(\psi_{c,m}(X_1, \ldots, X_c)) \\
H_{n,m}^{(c)} &:= \frac{1}{\binom{n}{c}} \sum_{j \in \mathcal{A}_{n,c}} h_m^{(c)}(X_{j_1}, \ldots, X_{j_c})
\end{aligned}$$

Note that the $H_m^{(c)}$ are U-statistics of degree c and mutually orthogonal. This leads to the well-known H-decomposition:

$$U_{n,m} = \sum_{c=1}^{m} \binom{m}{c} H_{n,m}^{(c)}$$

The first lemma considers asymptotic normality in the case that $m/n \to 0$.

Lemma 11. *Let $X_1, \ldots, X_n$ be i.i.d., and let $h_m : \mathbb{R}^m \to \mathbb{R}$ be a symmetric function, with $\mathbb{E}\, h_m(X_1, \ldots, X_m)^2$ uniformly integrable. Then, if*

$$\frac{m}{n} \frac{\sigma_{m,m}^2}{m\sigma_{1,m}^2} \to 0$$

then:

$$\frac{\sqrt{n} U_{n,m}}{m\sigma_{1,m}} \Rightarrow \mathcal{N}(0, 1)$$

Proof. As in Peng et al. (2019) and Minsker (2022), we write $U_{n,m}$ in terms of $H_{n,m}^{(1)}$ and a remainder term. Via the H-decomposition, we have that:

$$U_{n,m} = mH_{n,m}^{(1)} + R_{n,m}$$

Furthermore, $\mathrm{Var}(U_{n,m}) = m^2 \,\mathrm{Var}(H_{n,m}^{(1)}) + \mathrm{Var}(R_{n,m})$. We have expressions for the variance of $H_{n,m}^{(1)}$ and $\mathrm{Var}(U_{n,m})$ (see, for example, Lee (2019)):

$$\mathrm{Var}(R_{n,m}) = \sum_{k=2}^{m} \binom{m}{k}^2 \binom{n}{k}^{-1} \delta_{k,m}^2$$
$$\mathrm{Var}(H_{n,m}^{(1)}) = \frac{\sigma_{1,m}^2}{n}$$

Since $H_{n,m}^{(1)}$ is a sum of i.i.d. random variables with $m\sigma_{1,m}^2 < \sigma_{m,m}^2$, h_m^2 uniformly integrable, if we can show that $\mathrm{Var}(\sqrt{n}R_{n,m})/(m^2\sigma_{1,m}^2) \to 0$, we are done. By the H-decomposition and the orthogonality of the $H_{n,m}^{(c)}$, noting that $\delta_{1,m}^2 = \sigma_{1,m}^2$ and $\delta_{c,m}^2 \leq \binom{m}{k}^{-1}\sigma_{m,m}^2$, we have that:

$$\begin{aligned}
\mathrm{Var}(R_{n,m}) &= \sum_{k=2}^{m} \binom{m}{k}^2 \binom{n}{k}^{-1} \delta_{k,m}^2 \\
&\leq \sum_{k=2}^{m} \binom{m}{k} \binom{n}{k}^{-1} \sigma_{m,m}^2 \\
&= \sigma_{m,m}^2 \left(\sum_{k=0}^{m} \binom{m}{k} \binom{n}{k}^{-1} - \frac{m}{n} - 1 \right) \\
&= \sigma_{m,m}^2 \left(\frac{n+1}{n-m+1} - \frac{m}{n} - 1 \right) \\
&= \sigma_{m,m}^2 \frac{m-1}{n-m+1} \frac{m}{n}
\end{aligned}$$

where we used the combinatorial identity $\sum_{k=0}^{m} \binom{m}{k}\binom{n}{k} = \frac{n+1}{n-m+1}$. This means that if $(m/n)\sigma_{m,m}^2/(m\sigma_{1,m}^2) \to 0$, then:

$$\begin{aligned}
\frac{\mathrm{Var}(\sqrt{n}R_{n,m})}{m^2\sigma_{1,m}^2} &\leq \frac{m-1}{n-m+1} \frac{\sigma_{m,m}^2}{m\sigma_{1,m}^2} \\
&\to 0
\end{aligned}$$

□

The next lemma is mentioned in Peng et al. (2019); it is not necessary for our treatment, but we give a proof here for the curious reader:

Lemma 12. *Let $X_1, \ldots, X_n$ be i.i.d., and let $h_m : \mathbb{R}^m \to \mathbb{R}$ be a symmetric function, with*

$\mathbb{E}\, h_m(X_1, \ldots, X_m)^2$ *uniformly integrable. Then, if*

$$\frac{\sigma^2_{m,m}}{m\sigma^2_{1,m}} \to 1$$

then:

$$\frac{\sqrt{n}U_{n,m}}{m\sigma_{1,m}} \Rightarrow \mathcal{N}(0,1)$$

Proof. Similar to the proof of Lemma 11, we need to show $\mathrm{Var}(\sqrt{n}R_n)/m^2\sigma^2_{1,m} \to 0$. We will use the following facts from, e.g., Lee (2019):

$$\sigma^2_{k,m} \le \frac{k}{m}\sigma^2_{m,m},\ \forall k \le m \tag{33}$$

$$\sum_{k=1}^{m} \binom{n}{m}^{-1}\binom{m}{k}\binom{n-m}{m-k}k = \frac{m^2}{n} \tag{34}$$

As an alternative expression for $\mathrm{Var}(R_{n,m})$, we can exploit the orthogonality of $R_{n,m}$ and $H^{(1)}_{n,m}$:

$$\begin{aligned}
\mathrm{Var}(R_{n,m}) &= \mathrm{Var}(U_{n,m}) - m^2\,\mathrm{Var}(H^{(1)}_{n,m}) \\
&= \binom{n}{m}^{-1}\sum_{k=1}^{m}\binom{m}{k}\binom{n-m}{m-k}\sigma^2_{k,m} - \frac{m^2}{n}\sigma^2_{1,m} \\
&\le \frac{\sigma^2_{m,m}}{m}\binom{n}{m}^{-1}\sum_{k=1}^{m}\binom{m}{k}\binom{n-m}{m-k}k - \frac{m^2}{n}\sigma^2_{1,m} \\
&= \frac{m}{n}\sigma^2_{m,m} - \frac{m^2}{n}\sigma^2_{1,m} \\
&= \frac{m^2}{n}\sigma^2_{1,m}\left(\frac{\sigma^2_{m,m}}{m\sigma^2_{1,m}} - 1\right)
\end{aligned}$$

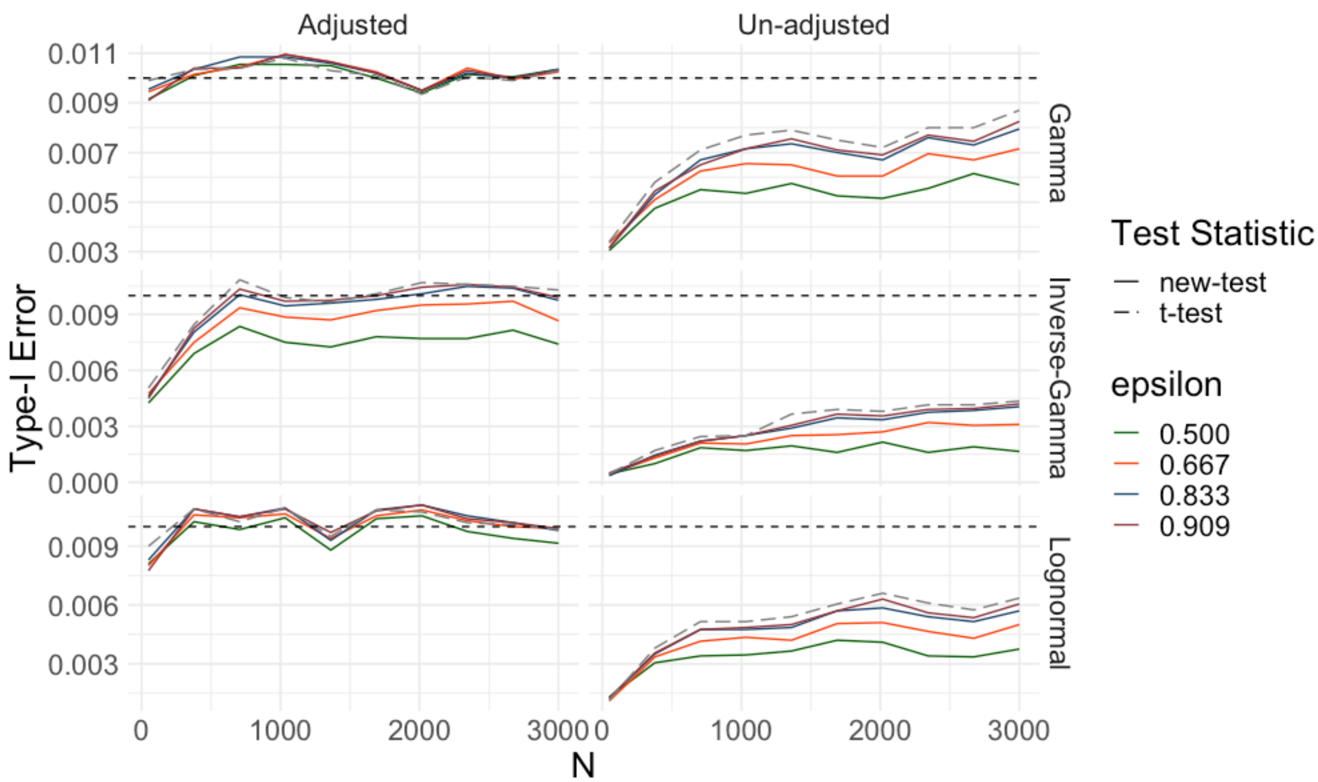
Thus, we have that

$$\frac{\mathrm{Var}(\sqrt{n}R_{n,m})}{m^2\sigma^2_{1,m}} \le \frac{\sigma^2_{m,m}}{m\sigma^2_{1,m}} - 1$$

and the result follows. □

C Size simulations

In this section we report some results on the size of the test, and how it compares with the classic t-test. For these simulations, we ran a simple one-sided t-test: $H_0 : \mu = 1$ against a 1-sided alternative. We used three different distributions for the data: gamma, lognormal, and inverse-gamma. All we normalized so that the mean was 1 and the variance was 1 as well. For the new test, we set epsilon a bit on the higher end: $\{0.500, 0.667, 0.833, 0.909\}$, where $m = an^{\epsilon}$, and again a is set so that m starts at 30. We also compared results when we used the monotone cubic transform of Hall (1992). For the new test, the adjustment is performed on the subsamples, and then the aggregation procedure is performed.



The right panel shows the un-adjusted test-statistics. Notice that none of the test-statistics achieve good size control, and when m is particularly slow-growing, the distortion is quite bad. On the other hand, for moderate choices of ϵ, when the adjustment is performed, the new test controls size just as well as the t-statistic. These simulations demonstrate that unlike classic robust procedures, which may not be consistent for the mean, the new proposed procedure is.