Simple Le Cam optimal inference for the tail weight of multivariate Student $t$ distributions: testing procedures and estimation

Christophe Ley$^{a,*}$, Anouk Neven$^{b}$

$^{a}$Université libre de Bruxelles, Département de Mathématique, Campus Plaine, boulevard du Triomphe, CP210, B-1050, Brussels, Belgium

$^{b}$Université du Luxembourg, UR en mathématiques, Campus Kirchberg, 6, rue Richard Coudenhove-Kalergi, L-1359, Luxembourg, Grand-Duchy of Luxembourg

Abstract

The multivariate Student $t$ distribution is at the core of classical statistical inference and is also a well-known model for empirical financial data. In the present paper, we propose optimal (in the Le Cam sense) inferential procedures about its tail weight parameter $\nu$. We start by establishing the uniform local asymptotic normality (ULAN) property of the multivariate location-scatter-tail weight Student $t$ model, which happens to be a non-trivial result. The ULAN structure then enables us to derive locally and asymptotically optimal (in the maximin sense) tests for tail weight under unspecified location and scatter. The Le Cam approach permits to replace these unknown quantities by any root-$n$ consistent estimators. The resulting tests thus improve on the classical approaches (likelihood ratio test, Wald test, Rao score test) by their flexibility and simplicity; moreover, we can write out explicitly the power of our tests against sequences of contiguous local alternatives. Regarding tail weight estimators, the one-step estimation procedure inherent to the ULAN property allows us to turn existing root-$n$ consistent estimators into fully efficient ones (in particular, we render the Mardia estimator optimal under multivariate Student distributions with $\nu > 8$). The finite-sample properties of our tests and estimators are analyzed in a large Monte Carlo simulation study, and we finally apply our methods on a financial data set.

AMS 2000 subject classifications: Primary 62F03; Secondary 62H15, 62F12

Keywords and Phrases: local asymptotic normality; locally asymptotically maximin tests; one-step estimation; Student $t$ distribution; tail weight

1. Introduction

1.1. Student $t$ distribution

Though already introduced by Helmert (1875), Lüroth (1876) and Pearson (1895), the univariate $t$ distribution was rediscovered by William Sealy Gosset, who published under the pseudonym Student on arrangement with his employers (see Student 1908), whence the commonly used expression Student $t$ on arrangement with his employers (see Student 1908), whence the commonly used expression Student $t$ distribution. In his work, he developed a small-sample theory and studied the distribution of

\[
 z := \frac{\bar{X} - \mu}{s} := \frac{1}{\sqrt{n-1}} \frac{\sum_{i=1}^{n} X_i - \mu}{\sum_{i=1}^{n} (X_i - \bar{X})^2},
\]

*Corresponding author, phone number: +32 2 650 58 78

Email addresses: chrisley@ulb.ac.be (Christophe Ley), anouk.neven@uni.lu (Anouk Neven)
where \( X_1, \ldots, X_n \) is a random sample from a normal population with unknown mean \( \mu \) and unknown variance \( \sigma^2 \). He showed that the hitherto assumed normal approximation may be very bad if the sample size is small. In a breakthrough paper published in 1925, Sir Ronald A. Fisher shifted the focus from \( z := \sqrt{n-1} z \) : he derived in a rigorous and elegant proof the distribution of \( t \) and coined the terminology “t distribution” (see Fisher 1925 or the historical paper Box 1981 relating the correspondence between Fisher and “Student”). Fisher’s paper and the related early success motivated researchers to generalize the \( t \) distribution to higher dimensions; the resulting multivariate \( t \) distribution has been studied, inter alia, by Cornish (1954) and Dunnett and Sobel (1954). Introduced originally as a normal-based sampling distribution, the Student \( t \) distribution is nowadays at the core of statistical inference, including e.g. robust estimation and hypothesis testing. Apart from its importance in classical inferential theory, application of the multivariate \( t \) distribution is a very promising approach in domains such as Bayesian analysis (Chien 2002) and cluster analysis (Peel and McLachlan 2000). We refer the reader to Kotz and Nadarajah (2004) for an extensive overview of its applications.

Under its most common form, the \( k \)-dimensional \( t \) distribution admits the density

\[
f_{\mu, \Sigma, \nu}(x) := c_{\nu,k} |\Sigma|^{-1/2} \left( 1 + \frac{1}{\nu} (x^\top \Sigma^{-1} x / (x^\top \Sigma^{-1} x)^{\nu/2}) \right)^{-\frac{k+\nu}{2}}, \quad x \in \mathbb{R}^k,
\]

with location parameter \( \mu \in \mathbb{R}^k \), scatter parameter \( \Sigma \in S_k \), the class of symmetric and positive definite \( k \times k \) matrices, and tail weight parameter \( \nu \in \mathbb{R}_+^+ \), and with normalizing constant

\[
c_{\nu,k} := \frac{\Gamma \left( \frac{\nu+k}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{k}{2} \right)},
\]

where the Gamma function is given by \( \Gamma(z) = \int_0^\infty \exp(-t)t^{z-1} dt \). Denoting by \( F_{\mu, \Sigma, \nu} \) the cumulative distribution function, we deduce from (1) that

\[
F_{\mu, \Sigma, \nu}(x) = O \left( \|x\|^{-\nu} \right)
\]

as \( \|x\| \to \infty \), which makes the Student \( t \) distribution a member of the class of heavy-tailed distributions with tail index \( \gamma = \frac{1}{2} \). It follows that the kurtosis is regulated by the parameter \( \nu \) : the smaller \( \nu \), the heavier the tails. For instance, for \( \nu = 1 \), we retrieve the fat-tailed Cauchy distribution, whereas, when \( \nu \) tends to infinity, we obtain the multivariate Gaussian distribution.

This heavy tail property is the key to the burgeoning popularity of the \( t \) distribution in empirical financial data modeling. Mandelbrot (1963) is the first to show that asset returns do not follow a Gaussian distribution, but have heavier tails. Ever since, there has been a great deal of empirical evidence supporting the existence of heavy-tailed models in finance and thereby challenging the well-established classical Gaussian assumption (see, e.g., Fama 1965 or Richardson and Smith 1993). Among these models, the class of \( t \) distributions has been suggested as a tractable, more viable alternative, particularly because it captures the observed fat tails (see, e.g., Blattberg and Gonedes 1974, Hagerman 1978, Perry 1983, Boothe and Glassman 1987 or Kan and Zhou 2003). For example, Kan and Zhou (2003) found out that the multivariate normality assumption on the distribution of Fama and French (1993)’s asset returns is rejected by a kurtosis test with a \( p \)-value of less than 0.01%. On the other hand, if the data is assumed to be drawn from a \( t \) distribution with \( \nu = 10 \) and 8, the kurtosis test has \( p \)-values of 0.34% and 16.84%, respectively. It follows that a \( t \) distribution with tail weight \( \nu = 10 \) is rejected by the data, whereas a value of \( \nu = 8 \) seems to be an adequate model. Hence, by advocating the use of the \( t \) distribution, a natural problem of interest consists in testing the null hypothesis \( H_0 : \nu = \nu_0 \) (where \( \nu_0 > 0 \) is fixed) against alternatives of the form \( H_{1+}^\nu : \nu \neq \nu_0 \) (two-sided test), \( H_{1-}^\nu : \nu < \nu_0 \) or \( H_{1+}^{-\nu} : \nu > \nu_0 \) (one-sided tests).
The likelihood ratio test (LR hereafter) provides the standard way to tackle this question. Clearly, the underlying test statistic invokes (i) the maximum likelihood estimators of the triple \((\mu, \Sigma, \nu)\) as solutions of the maximization of the log-likelihood function based on the \(t\) density without any constraint, and (ii) the maximum likelihood estimators of \((\mu, \Sigma)\) subject to the restriction \(\nu = \nu_0\). Unfortunately, there exists no closed-form solution to these maximization problems in the Student \(t\) family and hence numerical procedures are required. A standard approach for solving numerically the likelihood equations is the popular EM algorithm of Dempster et al. (1977) or some variants of it discussed for the Student \(t\) case in Liu and Rubin (1995). However, when \(\nu\) is small or unknown, Liu and Rubin (1995) have expressed the warning that ML procedures might be misled due to numerous spikes with very high likelihood mass in the likelihood function. Moreover, especially for higher dimensions, the convergence of ML algorithms can be very slow.

Common large sample alternatives to the likelihood ratio test are the Wald test (W) and the Rao score test (RS) (or Lagrange multiplier test). Similar to the likelihood ratio test statistic, the underlying test statistics are based on maximum likelihood estimators: the Wald test requires the computation of the maximum likelihood estimator of the triple \((\mu, \Sigma, \nu)\), whereas the Rao score test is derived from a constrained maximization problem, namely the maximization of the log-likelihood with respect to location and scatter subject to the constraint \(\nu = \nu_0\). The LR, W and RS tests, regarded as the Holy Trinity in asymptotic statistics, are known to share the same efficiency properties (see Engle 1984), but all three suffer from the non-existence of exact expressions and hence from numerical complexity in the (univariate and) multivariate \(t\) case (although this complexity is less heavy for the RS test).

In the present paper, we therefore propose a new technique for tackling hypothesis testing on the tail weight parameter \(\nu\) under unspecified location and scatter. Our testing procedures will be asymptotically as powerful as the LR, W and RS tests, but improve on the latter by their flexibility and simplicity. Moreover, as we shall see, our approach (described in Section 1.2 below) will allow us to write out explicitly the powers of our tests against sequences of contiguous alternatives, which in general is extremely difficult to achieve with the classical tests. Our technique also enables us to improve the power of existing root-\(n\) consistent estimators of the tail weight parameter; more precisely, they will become as powerful as the ML estimator. This is of particular interest for estimators having a closed-form expression, as they will then combine simplicity with efficiency.

1.2. Optimal parametric tests and estimators: the Le Cam methodology

As already mentioned above, the main purpose of the present work is to derive simple yet efficient tests for the tail weight parameter of Student \(t\) distributions, more precisely, tests that are locally and asymptotically optimal. The underpinning optimality in this paper is the so-called maximin optimality. Recall that a test \(\phi^*\) is called maximin in the class \(C_\alpha\) of level-\(\alpha\) tests for \(H_0\) against \(H_1\) if (i) \(\phi^*\) has level \(\alpha\) and (ii) the power of \(\phi^*\) is such that

\[
\inf_{P \in H_1} \mathbb{E}_P[\phi^*] \geq \sup_{\phi \in C_\alpha} \inf_{P \in H_1} \mathbb{E}_P[\phi].
\]

Our second goal in this paper consists in constructing optimal one-step estimators for tail weight, meaning that we start from a given root-\(n\) consistent estimator and render it as efficient as the ML estimator. The tools involved will not differ from those established for testing purposes. The most prominent example of estimator we shall improve is the celebrated Mardia (1970) estimator of tail weight.

The backbone of our constructions will be the famous “Le Cam methodology”. The concept of local asymptotic normality (LAN) is among Le Cam’s best-known contributions and plays an essential role in asymptotic optimality theory. To the best of our knowledge, nobody has yet taken advantage of the LAN
approach in the framework of tail parameter inference for univariate and multivariate $t$ distributions. In order to ease the reading, we will briefly review here the LAN property and its contribution to the theory of hypothesis testing and point estimation. The following definition of LAN corresponds to Le Cam and Yang (2000).

For all $n$, let $E^{(n)} = \left( \mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} := \{ P_{\theta}^{(n)} | \theta \in \Theta \subset \mathbb{R}^k \} \right)$ be a sequence of $\theta$-parametric statistical models, called experiments in Le Cam’s terminology, and let $\delta_n$ be a sequence of positive numbers going to zero. The family $\mathcal{P}^{(n)}$ is called LAN at $\theta \in \Theta$ if there exists a sequence of random vectors $\Delta^{(n)}(\theta)$, called central sequence, and a non-singular symmetric matrix $J(\theta)$, the associated Fisher information matrix, such that, for every bounded sequence of vectors $h_n \in \mathbb{R}^k$,

$$
\log \frac{dP_{\theta+\delta_n h_n}^{(n)}}{dP_{\theta}^{(n)}} - h_n' \Delta^{(n)}(\theta) + \frac{1}{2} h_n' J(\theta) h_n = o_P(1) \quad (2)
$$

and $\Delta^{(n)}(\theta) \overset{L}{\to} N_k(0, J(\theta))$, both under $P_{\theta}^{(n)}$ as $n \to \infty$.

We easily see that the log-likelihood ratios $\log \frac{dP_{\theta+\delta_n h_n}^{(n)}}{dP_{\theta}^{(n)}}$ in (2) of a LAN family behave asymptotically like the log-likelihood ratio of the classical Gaussian shift experiment

$$
\mathcal{E}_{J(\theta)} = (\mathbb{R}^k, B_k, P_{\theta} := \{ P_{h,\theta} = N_k(\theta, J(\theta), h) | h \in \mathbb{R}^k \})
$$

with a single observation which we denote as $\Delta$. This approximation of the statistical experiments $E^{(n)}$ by the normal experiment $\mathcal{E}_{J(\theta)}$ was Le Cam’s main motivation: the family of probability measures under study can be approximated very closely by a family of simpler nature (Le Cam 1960). It has, as we shall see, important consequences on the construction of locally and asymptotically optimal testing procedures and estimates.

In his 1964 paper, Le Cam formalizes his idea of “closely approximated” statistical models by introducing the deficiency distance $\Delta(\mathcal{E}, \mathcal{F})$ between two experiments $\mathcal{E}$ and $\mathcal{F}$. Roughly speaking, for bounded loss functions, the deficiency of $\mathcal{E}$ relative to $\mathcal{F}$ is small if and only if to every statistical procedure for $\mathcal{E}$ there exists a corresponding procedure for $\mathcal{F}$ with almost the same risk function. A few years later, Le Cam establishes the equivalence between the weak convergence in terms of the deficiency distance and the weak convergence of likelihood ratio processes. The latter connection constitutes the breakthrough of his work, as for every LAN model $E^{(n)}$,

$$
\Delta \left( E^{(n)}, \mathcal{E}_{J(\theta)} \right) \to 0
$$

when $n \to \infty$. In the context of hypothesis testing, this means that, asymptotically, all power functions that are implementable in the local experiments $E^{(n)}$ are the power functions that are possible in the Gaussian shift experiment $\mathcal{E}_{J(\theta)}$. In view of these considerations, it follows that asymptotically optimal tests in the local models can be derived by analyzing the Gaussian limit model. More precisely, if a test $\phi(\Delta)$ enjoys some exact optimality property in the Gaussian experiment $\mathcal{E}_{J(\theta)}$, then the corresponding sequence $\phi(\Delta^{(n)})$ inherits, locally and asymptotically, the same optimality properties in the sequence of experiments $E^{(n)}$.

A further attractive feature of the Le Cam methodology and, more precisely, of the LAN property lies in the construction of efficient estimators. In Le Cam (1986, Chapter 11), Le Cam makes use of the expansion (2) to develop the famous one-step estimation technique, transforming an arbitrary preliminary estimator $\hat{\theta}_n$ of $\theta$ (only optimal in terms of consistency rates) into an estimator $\tilde{\theta}_n$ satisfying

$$
\delta_n^{-1}(\hat{\theta}_n - \theta) = J(\theta)^{-1} \Delta^{(n)}(\theta) + o_P(1)
$$
under $P^{(n)}_{\theta}$ as $n \to \infty$. The resulting estimator $\tilde{\theta}_n$ is called best regular in the terminology of Hájek (1970), and is locally and asymptotically optimal in the sense that it attains efficiency asymptotically.

In the present work, we will investigate the LAN phenomenon for a sequence of (univariate and multivariate) Student $t$ distributions with respect to the location, scatter and tail parameters (more precisely, we shall obtain the uniform LAN property; see Section 2.2). As explained above, the (U)LAN property allows one to transfer the well-known optimal procedures from the classical Gaussian shift experiment to the $t$ model and thus paves the way towards the construction of locally and asymptotically optimal (in the maximin sense) tests for the tail parameter under unspecified location and scatter. As we shall see in the sequel, our test statistic strongly resembles that of the RS test; however, our approach allows us to use any root-$n$ consistent estimators of location and scatter under fixed $\nu = \nu_0$ (e.g., for $\nu_0 > 4$, we can even use the mean vector and variance-covariance matrix), hence circumvents their ML estimation. Our new tests, which are asymptotically equivalent to the Holy Trinity, will thus supersede the classical approaches thanks to their computational simplicity. Moreover, the Le Cam framework will allow us to provide an explicit formula for the power function of our tests. Regarding point estimation, we transform, among others, the root-$n$ consistent Mardia estimator into a locally and asymptotically optimal estimator.

1.3. Outline of the paper

The paper is organized as follows. In Section 2, we establish the main theoretical result of the paper, namely the uniform local asymptotic normality property (with respect to the location, scatter and tail parameters). In Section 3 we then construct the new optimal tests for tail weight. We first explain in Section 3.1 how to estimate the nuisance parameters and identify then in Section 3.2 the efficient central sequence for tail weight under unknown location and scatter. In Section 3.3 we derive, thanks to the ULAN property, the locally and asymptotically optimal tests. We study the asymptotic behavior of our test statistic, both under the null and under a sequence of local alternatives, allowing us to compute explicitly the power of our test. We devote Section 4 to a description of our one-step estimator for tail weight. Special attention is paid to the Mardia estimator (under multivariate Student distributions with $\nu > 8$) and to the McElroy estimator (under univariate $t$ distributions with small $\nu$). We explore in Section 5 the finite-sample properties of our proposed testing and estimation procedures via a large Monte Carlo simulation study. In Section 6, we apply our methods on a financial data set. Finally, the Appendix contains the proof of the ULAN property.

2. Uniform local asymptotic normality (ULAN)

Throughout, the data points $X_1, \ldots, X_n$ are assumed to follow a multivariate $t$ distribution with parameters $(\mu, \Sigma, \nu) =: \mathbf{v}$. The relevant statistical experiment thus involves the parametric family

$$P^{(n)}_{\mathbf{v}} = \{P^{(n)}_{\mathbf{v}} | \mathbf{v} \in \mathbb{R}^k \times S_k \times \mathbb{R}_0^+\},$$

where $P^{(n)}_{\mathbf{v}}$ stands for the joint distribution of $X_1, \ldots, X_n$. The rest of this section is devoted to the establishment of the crucial ULAN property of the considered parametric family. Indeed, as explained in the Introduction, this Taylor type expansion of the log-likelihood ratio is the backbone of our construction of optimal tests for the tail parameter $\nu$ under unspecified location $\mu$ and unspecified scatter $\Sigma$.

2.1. Notation and definitions

In order to ease readability, we start by explaining some notations that will be useful in the sequel. Set $M_k$ the class of $k \times k$ matrices. We write $\text{vec}(A)$ for the $k^2$-vector obtained by stacking the columns
of a matrix $A \in M_k$ on top of each other, and $\text{vech}(A)$ for the $k(k+1)/2$ subvector of $\text{vec}(A)$ where only the upper diagonal entries in $A$ are considered. Define $P_k$ as the $k(k+1)/2 \times k^2$ matrix such that $P_k \text{vech}(A) = \text{vec}(A)$ for any $k \times k$ symmetric matrix $A$. Denoting by $e_\ell$ the $\ell$th vector in the canonical basis of $\mathbb{R}^k$ and by $I_k$ the $k \times k$ identity matrix, let

$$K_k := \sum_{i,j=1}^k (e_i e'_j) \otimes (e_i e'_j) \quad \text{and} \quad J_k := \sum_{i,j=1}^k (e_i e'_j) \otimes (e_i e'_j) = \text{vec}(I_k)(\text{vec}(I_k))',$$

where the $k^2 \times k^2$ matrix $K_k$ is known as the commutation matrix. With this notation, $K_k(\text{vec}(A)) = \text{vec}(A')$ and $J_k(\text{vec}(A)) = (\text{tr}A)(\text{vec}(I_k))$. Finally, we write $A \otimes^2$ for the usual Kronecker product $A \otimes A$.

For multivariate $t$ distributions, we define the score vector $\mathbf{L}_V(x) := \left(\mathbf{L}_V^{(i)}(x)\right)_{i=1,2,3}$ for $V$ as

$$\mathbf{L}_V(x) := \begin{pmatrix} D_{\mathbf{\mu}} \log(f_{\mathbf{\mu},\Sigma,V}(x)) \\ D_{\text{vech}(\Sigma)} \log(f_{\mathbf{\mu},\Sigma,V}(x)) \\ D_{\nu} \log(f_{\mathbf{\mu},\Sigma,V}(x)) \end{pmatrix},$$

where $D_{\text{vech}(\Sigma)}$ and $D_{\mathbf{\mu}}$ are the usual gradients and $D_{\nu}$ is the usual derivative. Direct computations lead to

$$\mathbf{L}_V(x) = \begin{pmatrix} 1/2 P_k(\Sigma^{(2)})^{-1/2} \text{vec} \left( \frac{1+k/\nu}{1+\|\Sigma^{-1/2}(x-\mu)^2/\nu \Sigma^{-1}(x-\mu)} \right) \\ c_{\nu,k}' c_{\nu,k} \nu + 1 \nu/\nu - 1/2 \log \left( 1 + \|\Sigma^{-1/2}(x-\mu)^2/\nu \Sigma^{-1/2}(x-\mu)^2/\nu + \frac{\nu+k}{\nu} \|\Sigma^{-1/2}(x-\mu)^2/\nu \right) \right) \\ \end{pmatrix},$$

where $c_{\nu,k}'$ stands for the derivative of the mapping $\nu \mapsto c_{\nu,k}$. Now, note that the $\mu$-score is anti-symmetric in $x - \mu$, while the scores for $\Sigma$ and $\nu$ are symmetric in $x - \mu$. Hence, the symmetry properties with respect to $x - \mu$ entail that the resulting Fisher information matrix $\Gamma(V)$, given by the covariance matrix of the score vector, partitions into

$$\begin{pmatrix} \Gamma_{11}(V) & 0 & 0 \\ 0 & \Gamma_{22}(V) & \Gamma_{23}(V) \\ 0 & \Gamma_{23}(V)' & \Gamma_{33}(\nu) \end{pmatrix}. \quad (3)$$

Lange et al. (1989) derived explicit expressions for the entries of $\Gamma(V)$ given by

$$\Gamma_{11}(V) = \frac{\nu+k}{\nu+k+2} \Sigma^{-1},$$

$$\Gamma_{22}(V) = \frac{1}{4} P_k(\Sigma^{(2)})^{-1/2} \left[ \frac{\nu+k}{\nu+k+2} (I_k^2 + K_k + J_k) - J_k \right](\Sigma^{(2)})^{-1/2}P_k',$$

$$\Gamma_{23}(V) = \frac{1}{(\nu+k+2)(\nu+k)} P_k(\Sigma^{(2)})^{-1/2} \text{vec}(I_k),$$

and

$$\Gamma_{33}(\nu) = -\frac{1}{2} \left[ \frac{\nu+k}{2} \psi' \left( \frac{\nu+k}{2} \right) - \frac{1}{2} \psi' \left( \frac{\nu}{2} \right) + \frac{k}{\nu(\nu+k)} - \frac{1}{\nu+k} + \frac{\nu+2}{\nu(\nu+k+2)} \right],$$

where $\psi'$ is the trigamma function. Ley and Paindaveine (2010) recently proved that the information matrix is finite and non-singular for all $V \in \mathbb{R}^k \times S_k \times \mathbb{R}_{++}^*$.

Finally, for two parameter sets $V_1 = (\mu_1, \Sigma_1, \nu_1)$ and $V_2 = (\mu_2, \Sigma_2, \nu_2)$, we will make throughout the slight abuse of notation and write $V_1 + V_2$ for $(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2, \nu_1 + \nu_2)$. 
2.2. Uniform local asymptotic normality (ULAN)

With these notations and definitions in hand, we are ready to state the main technical result of this paper, namely the announced ULAN property of the multivariate $t$ family with respect to the location, scatter and tail parameters.

**Theorem 1.** For any $\mu \in \mathbb{R}^k$, $\Sigma \in S_k$ and $\nu > 0$, the multivariate Student $t$ family $P_\nu^{(n)}$ is ULAN at $\nu$, with central sequence

$$\Delta^{(n)}(\nu) := \left(\Delta^{(n)}_{i}(\nu)\right)_{i=1,2,3} = \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} L^{(i)}_{\nu}(X_k)\right)_{i=1,2,3},$$

and information matrix $\Gamma(\nu)$. More precisely, for any $\nu^{(n)} = (\mu^{(n)}, \Sigma^{(n)}, \nu^{(n)}) = \nu + O(n^{-1/2})$ and for any bounded sequence $\nu^{(n)} = (\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)}) \in \mathbb{R}^k \times M_k \times \mathbb{R}$ such that $\Sigma^{(n)} + n^{-1/2} \tau_2^{(n)} \in S_k$ and $\nu^{(n)} + n^{-1/2} \tau_3^{(n)} > 0$, we have

$$\Lambda^{(n)}_{\nu^{(n)} + n^{-1/2} \tau^{(n)}}(X_1, \ldots, X_n) := \frac{\log \left(\frac{dP^{(n)}_{\nu^{(n)} + n^{-1/2} \tau^{(n)}}}{dP^{(n)}_{\nu^{(n)}}}\right)}{\tau^{(n)}} = \frac{1}{2} \left((T^{(n)})' \Delta^{(n)} \left(\nu^{(n)}\right) - \frac{1}{2} (T^{(n)})' \Gamma(\nu) T^{(n)} + o_P(1)\right),$$

where $T^{(n)} := (\mu^{(n)}', (\text{vech}(\Sigma^{(n)}))', \tau_3^{(n)}')$, and $\Delta^{(n)}(\nu^{(n)}) \xrightarrow{L} N_{k+k(k+1)/2+1}(0, \Gamma(\nu))$, both under $P^{(n)}_{\nu^{(n)}}$, as $n \to \infty$.

**Proof.** Our proof of this ULAN result relies on Lemma 1 from Swensen (1985) – more precisely on its extension by Garel and Hallin (1995). The sufficient conditions for ULAN given in Swensen’s result follow from standard arguments (hence are left to the reader), once it is shown that $(\mu', (\text{vech}(\Sigma))', \nu') \mapsto f_{\nu^{1/2} \Sigma^{1/2}, \nu}^1(x)$ is differentiable in quadratic mean, where $f_{\mu, \Sigma, \nu}$ is the density in (1). The latter differentiability in quadratic mean spells out as

$$\int_{\mathbb{R}^k} \left\{ \frac{1}{\mu + s, \Sigma + H, \nu + t} (x) - \frac{1}{\mu, \Sigma, \nu} (x) - \frac{s}{\text{vech}(H)} t \delta \right\}^2 \left( \frac{D_{\mu, \Sigma, \nu} f_{\nu^{1/2} \Sigma^{1/2}, \nu}^1(x)}{D_{\mu, \Sigma, \nu} f_{\nu^{1/2} \Sigma^{1/2}, \nu}^1(x)} \right) \, dx = o \left( \left\| \frac{s}{\text{vech}(H)} t \right\|^2 \right)$$

for $s \in \mathbb{R}^k$, $H \in M_k$ and $t \in \mathbb{R}$ such that $\Sigma + H \in S_k$ and $\nu + t > 0$. This result is established in the Appendix (see Lemma 2) in a rather long and technical proof, which requires the use of two auxiliary lemmas, the concept of quadratic mean continuity and Lebesgue’s dominated convergence theorem. \hfill \Box

Note that the term “uniform” in ULAN indicates that the local asymptotic normality property holds not only at $\nu$, but in a neighborhood of that point. Further, the ULAN structure entails the following asymptotic linearity property of the central sequence:

$$\Delta^{(n)} \left(\nu + n^{-1/2} \tau^{(n)}\right) = \Delta^{(n)}(\nu) - \Gamma(\nu) T^{(n)} + o_P(1), \quad (4)$$

as $n \to \infty$, under $P^{(n)}_{\nu^{(n)}}$. As we shall see in the next section, the preceding asymptotic linearity property is needed, first to construct the optimal tests for tail weight under unspecified location and scatter and second to derive explicit expressions of the power function for those tests.
3. Locally and asymptotically optimal tests for tail weight

In this section, we make use of the ULAN property established in the previous section in order to construct locally and asymptotically optimal tests for the tail parameter $\nu$. To this end, we apply the “Le Cam methodology” for parametric tests to the present context. We shall proceed in three steps: first, we explain how to estimate the nuisance parameters $\mu$ and $\Sigma$ (Section 3.1), second we show how to modify the central sequence $\Delta^{(n)}$ in order to take into account the asymptotic correlation between the scatter and tail parameters (Section 3.2), and finally we write out the test statistic and study its asymptotic properties (Section 3.3).

3.1. Estimation of the nuisance parameters

Clearly, it is hard to think of any practical problem where the location and scatter are specified. We thus concentrate on asymptotic optimality under unspecified location and scatter: $\mu$ and $\Sigma$ play the roles of nuisance parameters, whereas $\nu$ is the parameter of interest. In particular, the central sequence $\Delta^{(n)}(Y^{(n)})$ given in Theorem 1 (or a modified version of it, see the end of Section 3.2) depends on the unknown values of $\mu$ and $\Sigma$, hence is not yet a true statistic. This problem can be solved by replacing $(\mu, \Sigma)$ with an adequate estimator $(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)})$ whilst, of course, paying attention to the asymptotic effects of such a substitution. It is precisely here that the asymptotic linearity property comes in handy. Of course, our aim consists in using relation (4) with $\tau^{(n)} = \left( n^{1/2}(\hat{\mu}^{(n)} - \mu), n^{1/2}(\hat{\Sigma}^{(n)} - \Sigma), 0 \right)$, providing the asymptotic link between $\Delta^{(n)}(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}; \nu)$ and $\Delta^{(n)}(\mu, \Sigma, \nu)$. However, this replacement is not straightforward and imposes one more condition on the estimators, which is summarized in the following assumption which we state for a general parametric model $P^{(n)}_\lambda$ (see Kreiss 1987, where such replacements have been worked out in detail).

**Assumption A.** The sequence of estimators $\hat{\lambda}^{(n)}$ defined for a sequence of experiments $P^{(n)}_\lambda = \{P^{(n)}_\lambda | \lambda \in \Lambda\}$ indexed by a parameter $\lambda$ belonging to the parameter space $\Lambda$ is

(i) root-$n$ consistent; that is, $n^{1/2}(\hat{\lambda}^{(n)} - \lambda) = O_p(1)$ under $P^{(n)}_\lambda$ for all $\lambda \in \Lambda$;

(ii) locally and asymptotically discrete; that is, the number of possible values of $\hat{\lambda}^{(n)}$ in $\lambda$-centered balls with $O(n^{-1/2})$ radius is uniformly bounded, as $n \to \infty$.

Both estimators $\hat{\mu}^{(n)}$ and $\hat{\Sigma}^{(n)}$ will have to satisfy this requirement in what follows. Local asymptotic discreteness is a concept that goes back to Le Cam (1986) and is quite standard in parameter estimation, since it turns root-$n$ consistent estimators into uniformly root-$n$ consistent ones (see Lemma 4.4 in Kreiss 1987). Denoting by $[x]$ the smallest integer larger than or equal to $x$ and by $c_0$ an arbitrary positive constant that does not depend on $n$, any sequence of estimators $\hat{\lambda}^{(n)}$ of $\lambda$ can be discretized by replacing it with

$$\hat{\lambda}^{(n)}_d := c_0^{-1} n^{-1/2} \text{sign} \left( \hat{\lambda}^{(n)} \right) \left[ c_0 n^{1/2} \hat{\lambda}^{(n)} \right].$$

In practice, however, such discretization is not required, as $c_0$ can be chosen large enough to make discretization be irrelevant at the fixed sample size $n$. Assumption A(ii) is thus a purely technical requirement with little practical implications, so that the preliminary estimator essentially only needs to be consistent at the standard root-$n$ rate. Obvious examples of such estimators for $\mu$ and $\Sigma$ are of course the sample mean $\bar{X}$ and (a multiple of) the sample covariance matrix $S(1 - 2/\nu_0)$ (with $\nu_0$ the value under the null). However, it is well-known that the latter require first respectively second moments in order to exist, and second respectively fourth moments in order to be root-$n$ consistent. In the univariate
case, simple moment-free estimators for location and scale are the median and the median of absolute deviations (MAD). In higher dimensions, if moment conditions are to be avoided, one can for example use as \( \hat{\mu}^{(n)} \) the spatial median of Möttönen and Oja (1995) and as \( \hat{\Sigma}^{(n)} \) Tyler (1987)’s shape estimator (adjusted to be a scatter estimator by multiplication of a scale estimator), or construct location and scatter estimators via the Minimum Covariance Determinant (MCD) method, the root-\( n \) consistency of which can be found in Cator and Lopuha¨a (2012). This underlines one of the advantages of our proposal: one can freely choose, according to the needs in a given situation, any of the above estimators (or others), and is not forced to use the more complicated maximum likelihood estimators for location and scatter.

### 3.2. An efficient central sequence for tail weight

The block-diagonal structure of the Fisher information matrix \( \mathbb{I} \) confirms that the blocks \( \mu \) and \( \Sigma, \nu \) are asymptotically uncorrelated and thus the non-specification of \( \mu \) does not affect, asymptotically, inferential procedures for \( \nu \) and/or \( \Sigma \). More precisely, it follows from the ULAN property that replacing \( \mu \) with an estimator \( \hat{\mu}^{(n)} \) satisfying Assumption A has no influence, asymptotically, on \( \Delta_{2}^{(n)}(\nu) \), the \( \nu \)-part of the central sequence. This can be seen via the asymptotic linearity (4) of the central sequence combined with Lemma 4.4 in Kreiss (1987) (if unclear, see the proof of Lemma 1, where we develop this argument). On the contrary, a non-zero asymptotic covariance \( \Gamma_{23}(\nu) \) between the scatter and the tail part of the central sequence implies that the cost of not knowing the actual value of \( \Sigma \) is strictly positive when performing inference on \( \nu \). This means that a local perturbation of \( \Sigma \) has the same asymptotic impact on \( \Delta_{2}^{(n)}(\nu) \) as a local perturbation of \( \nu \). This impact will be taken into account in what follows.

The ULAN structure and the convergence of the local experiments to the Gaussian shift experiment

\[
\begin{align*}
\left( \frac{\Delta_2}{\Delta_3} \right) \sim N_{k(k+1)/2+1} \left( \begin{pmatrix} \Gamma_{22}(\nu) & \Gamma_{23}(\nu) \\ \Gamma_{23}(\nu)' & \Gamma_{33}(\nu) \end{pmatrix} \right) \left( \begin{pmatrix} \text{vech}(\tau_2) \\ \tau_3 \end{pmatrix} \right),
\end{align*}
\]

where \( (\tau_2, \tau_3) \in M_k \times \mathbb{R} \), imply that locally and asymptotically optimal inference on \( \nu \) should be based on the residual of the regression of \( \Delta_3 \) with respect to \( \Delta_2 \), computed at \( \Delta_{2}^{(n)}(\nu) \) and \( \Delta_{2}^{(n)}(\nu) \). The resulting efficient central sequence for \( \nu \) thus takes the form

\[
\Delta_{2}^{(n)}(\nu) := \Delta_{2}^{(n)}(\nu) - (\Gamma_{23}(\nu))'(\Gamma_{22}(\nu))^{-1} \Delta_{2}^{(n)}(\nu).
\]

The projection of \( \Delta_{2}^{(n)}(\nu) \) onto the subspace orthogonal to \( \Delta_{2}^{(n)}(\nu) \) ensures that the new efficient central sequence is asymptotically uncorrelated with the central sequences corresponding to \( \mu \) and \( \Sigma \). Under \( \mathcal{B}(\nu) \), the efficient central sequence for tail weight is asymptotically normal, with mean zero and covariance (efficient Fisher information for tail weight)

\[
\Gamma_{33}^*(\nu) := \Gamma_{33}(\nu) - (\Gamma_{23}(\nu))'(\Gamma_{22}(\nu))^{-1} \Gamma_{23}(\nu),
\]

which is non-zero (see Theorem 4.2 from Ley and Paindaveine 2010). In the one-dimensional case, the efficient central sequence and Fisher information for \( \nu \) reduce after some elementary calculations to

\[
\Delta_{3}^{(n)}(\nu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\nu + 3}{2\nu^2} \left( \frac{X_i - \mu}{\sigma} \right)^2 \left( 1 + \frac{(X_i - \mu)^2}{\sigma^2\nu} \right)^{-1} - \frac{1}{2} \log \left( 1 + \frac{(X_i - \mu)^2}{\sigma^2\nu} \right) - \frac{c_\nu}{c_\nu} - \frac{1}{\nu(\nu + 1)} \right\}
\]

and

\[
\Gamma_{33}^*(\nu) = -\frac{1}{2} \left[ \frac{1}{2} \psi'(\frac{\nu + 1}{2}) - \frac{1}{2} \psi'(\frac{\nu}{2}) + \frac{\nu + 3}{\nu(\nu + 1)^2} \right],
\]

where \( \sigma^2 := \Sigma \). In particular, if \( k = 1 \), \( \Gamma_{33}^*(\nu) \) does not depend on the location nor on the scale (whence the notation). In the following result, we derive the asymptotic linearity property of the efficient central sequence for tail weight.
Lemma 1. For any $\mu \in \mathbb{R}^k$, $\Sigma \in S_k$ and $\nu > 0$, and for any bounded sequence $\tau^{(n)} = \left(\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)}\right) \in \mathbb{R}^k \times M_k \times \mathbb{R}$ such that $\Sigma + n^{-1/2} \tau_2^{(n)} \in S_k$ and $\nu + n^{-1/2} \tau_3^{(n)} > 0$, we have that, under $P_{\nu}^{(n)}$ and as $n \to \infty$,

$$
\Delta_3^{(n)*} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right) = \Delta_3^{(n)*} (\mathbf{V}) - \Gamma_{33}^{*} (\mathbf{V}) \tau_3^{(n)} + o_P(1).
$$

In particular, if $\tau_3^{(n)} = 0$,

$$
\Delta_3^{(n)*} \left( \mu + n^{-1/2} \tau_1^{(n)}, \Sigma + n^{-1/2} \tau_2^{(n)}, \nu \right) = \Delta_3^{(n)*} (\mu, \Sigma, \nu) + o_P(1).
$$

Proof. It follows from the asymptotic linearity property of the central sequence given in (4) that, under $P_{\nu}^{(n)}$ and for $n \to \infty$,

$$
\Delta_3^{(n)} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right) = \Delta_3^{(n)} (\mathbf{V}) - (\Gamma_{23}(\mathbf{V}))' \tau_2^{(n)} - \Gamma_{33}(\nu) \tau_3^{(n)} + o_P(1),
$$

and

$$
\Delta_2^{(n)} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right) = \Delta_2^{(n)} (\mathbf{V}) - \Gamma_{22}(\mathbf{V}) \tau_2^{(n)} - \Gamma_{23}(\mathbf{V}) \tau_3^{(n)} + o_P(1).
$$

By the definition of the central sequence for tail weight, we have

$$
\Delta_3^{(n)*} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right) = \Delta_3^{(n)} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right) - \left( \Gamma_{23}(\mathbf{V}) \right)' \left( \Gamma_{22}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)^{-1} \left( \Gamma_{22}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)^{-1} \Delta_2^{(n)} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right).
$$

Substituting (5) and (6) in (7) yields

$$
\Delta_3^{(n)*} \left( \mathbf{V} + n^{-1/2} \tau^{(n)} \right) = \Delta_3^{(n)} (\mathbf{V}) - \left[ \left( \Gamma_{23}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)' \left( \Gamma_{22}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)^{-1} \left( \Gamma_{23}(\mathbf{V}) \right)' \left( \Gamma_{22}(\mathbf{V}) \right)^{-1} \right] \Delta_2^{(n)} (\mathbf{V})
$$

$$
= \Delta_3^{(n)*} (\mathbf{V}) - \left( \Gamma_{23}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)' \left( \Gamma_{22}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)^{-1} \left( \Gamma_{23}(\mathbf{V}) \right)' \left( \Gamma_{22}(\mathbf{V}) \right)^{-1} \Delta_2^{(n)} (\mathbf{V})
$$

$$
- \left( \Gamma_{23}(\mathbf{V}) \right)' \tau_2^{(n)} - \Gamma_{33}(\nu) \tau_3^{(n)} + \left( \Gamma_{23}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)' \left( \Gamma_{22}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)^{-1} \Gamma_{23}(\mathbf{V}) \tau_2^{(n)}
$$

$$
+ \left( \Gamma_{23}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)' \left( \Gamma_{22}(\mathbf{V} + n^{-1/2} \tau^{(n)}) \right)^{-1} \Gamma_{23}(\mathbf{V}) \tau_3^{(n)} + o_P(1),
$$

under $P_{\nu}^{(n)}$ as $n \to \infty$. Hence, the result follows from the continuity of $\mathbf{V} \mapsto \Gamma(\mathbf{V})$ and the boundedness in probability of the central sequence. 

It is precisely here that Assumption A (and especially the local asymptotic discreteness) comes into play: for any estimators $\hat{\mu}^{(n)}$ and $\hat{\Sigma}^{(n)}$ satisfying Assumption A, Lemma 4.4 in Kreiss (1987) ensures that the asymptotic linearity property of Lemma 1 holds after replacement of $\left( \tau_1^{(n)}, \tau_2^{(n)}, 0 \right)$ by the random quantity $\left( n^{1/2} (\hat{\mu}^{(n)} - \mu), n^{1/2} (\hat{\Sigma}^{(n)} - \Sigma), 0 \right)$, which entails that, asymptotically under $P_{\nu}^{(n)}$,

$$
\Delta_3^{(n)*} \left( \hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, \nu \right) = \Delta_3^{(n)*} (\mu, \Sigma, \nu) + o_P(1).
$$

The latter (asymptotic) equality in probability will allow us to derive the asymptotic behavior of our optimal tests for tail weight in the next section.
3.3. Simple optimal tests for tail weight

As described in the Introduction, the ULAN property allows us to translate optimal procedures from Gaussian shift experiments into our Student $t$ model. This, in combination with the developments of the previous section, entails that the optimal test $\phi_{v_0}^{(n)}$ for $H_0 : \nu = \nu_0$ (with $\nu_0 > 0$ fixed) in the Student $t$ family with unspecified location $\mu$ and scatter $\Sigma$ should be based on the efficient central sequence for tail weight. More concretely, $\phi_{v_0}^{(n)}$ rejects the null (at asymptotic level $\alpha$) in favor of $H_1^\neq : \nu \neq \nu_0$ whenever the test statistic $|Q_{v_0}^{(n)}|$, with

$$Q_{v_0}^{(n)} := \frac{\Delta_3^{(n)*}(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, v_0)}{\sqrt{\Gamma_{33}^{(n)}(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, v_0)}}$$

where the estimators $\hat{\mu}^{(n)}$ and $\hat{\Sigma}^{(n)}$ satisfy Assumption A, exceeds $z_{\alpha/2}$, the $\alpha/2$-upper quantile of a standard Gaussian distribution. Thanks to (8), we can derive the asymptotic properties of $Q_{v_0}^{(n)}$, and hence also of $\phi_{v_0}^{(n)}$, in the next theorem.

**Theorem 2.** Fix $\nu_0 > 0$ and suppose that $\hat{\mu}^{(n)}$ and $\hat{\Sigma}^{(n)}$ satisfy Assumption A. Then

(i) $Q_{v_0}^{(n)}$ is asymptotically standard normal under $\cup_{\mu \in \mathbb{R}^k} \cup_{\Sigma \in \mathcal{S}_k} P_{(\mu, \Sigma, v_0)}^{(n)}$;

(ii) $Q_{v_0}^{(n)}$ is asymptotically normal with mean $\tau_3 \sqrt{\Gamma_{33}^{(n)}(\mu, \Sigma, v_0)}$ and variance 1 under $\cup_{\mu \in \mathbb{R}^k} \cup_{\Sigma \in \mathcal{S}_k} P_{(\mu, \Sigma, v_0)}^{(n)}$ and is locally and asymptotically maxmin for testing $H_0$ against $H_1^\neq$.

Proof. Since $\Gamma_{33}^{(n)}(\mathcal{V})$ is continuous in both $\mu$ and $\Sigma$, we readily have for any bounded sequence $(\tau_1^{(n)}, \tau_2^{(n)}) \in \mathbb{R}^k \times M_k$ that $\lim_{n \to \infty} \Gamma_{33}^{(n)}(\mu + n^{-1/2} \tau_1^{(n)}, \Sigma + n^{-1/2} \tau_2^{(n)}, v_0) = \Gamma_{33}^{(n)}(\mu, \Sigma, v_0)$. Since this convergence of course implies convergence in probability, Lemma 4.4 in Kriess (1987) allows us to replace the non-random quantities with root-$n$ consistent and locally and asymptotically discrete estimators. Hence, Slutsky’s Lemma, combined with Lemma 1 entails that under $P_{(\mu, \Sigma, v_0)}^{(n)}$,

$$Q_{v_0}^{(n)} = \frac{\Delta_3^{(n)*}(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, v_0)}{\sqrt{\Gamma_{33}^{(n)}(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, v_0)}} = \frac{\Delta_3^{(n)*}(\mu, \Sigma, v_0)}{\sqrt{\Gamma_{33}^{(n)}(\mu, \Sigma, v_0)}} + oP(1) \quad (9)$$

as $n \to \infty$. The proof of the statement in Part (i) then follows, since $\Delta_3^{(n)*}(\mu, \Sigma, v_0)$ is asymptotically $\mathcal{N}(0, \Gamma_{33}^{(n)}(\mu, \Sigma, v_0))$ under $\cup_{\mu \in \mathbb{R}^k} \cup_{\Sigma \in \mathcal{S}_k} P_{(\mu, \Sigma, v_0)}^{(n)}$ by the central limit theorem. Moreover, still under $P_{(\mu, \Sigma, v_0)}^{(n)}$ and for any bounded sequence $\tau^{(n)} = (\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)}) \in \mathbb{R}^k \times M_k \times \mathbb{R}$, we see that, as $n \to \infty$,

$$\left(\Lambda_{(\mu, \Sigma, v_0)}^{(n)} + n^{-1/2} \tau^{(n)}/(\mu, \Sigma, v_0)\right) \overset{\mathcal{L}}{\to} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ -\frac{1}{2} \tau \tau^T(\mu, \Sigma, v_0) \end{pmatrix}, \begin{pmatrix} \Gamma_{33}^{(n)}(\mu, \Sigma, v_0) & \tau_3 \Gamma_{33}^{(n)}(\mu, \Sigma, v_0) \\ \tau_3 \Gamma_{33}^{(n)}(\mu, \Sigma, v_0) & \tau_3 \Gamma_{33}^{(n)}(\mu, \Sigma, v_0) \end{pmatrix}\right),$$

where $\Lambda_{(\mu, \Sigma, v_0)}^{(n)} + n^{-1/2} \tau^{(n)}/(\mu, \Sigma, v_0)$ is the log-likelihood ratio and $\tau^{(n)} := ((\tau_1^{(n)})', (v(\tau_2^{(n)}))', \tau_3^{(n)})'$, with $\tau = \lim_{n \to \infty} \tau^{(n)}$. Le Cam’s third lemma thus implies that $\Delta_3^{(n)*}(\mu, \Sigma, v_0)$ is asymptotically
\( \mathcal{N}(\tau_3 \Gamma^*_{33}(\mu, \Sigma, \nu_0), \Gamma^*_{33}(\mu, \Sigma, \nu_0)) \) under \( \bigcup_{\mu \in \mathbb{R}^k} \bigcup_{\Sigma \in S_k} P^{(n)}_{(\mu, \Sigma, \nu_0 + n^{-1/2} \tau_3^{(n)})} \). Since (9) holds as well under \( P^{(n)}_{(\mu, \Sigma, \nu_0 + n^{-1/2} \tau_3^{(n)})} \) by contiguity, Part (ii) of the theorem readily follows.

As regards Part (iii), the fact that \( \phi^{(n)}_{\nu_0} \) has asymptotic level \( \alpha \) follows directly from the asymptotic null distribution given in Part (i), while local asymptotic maximinity is a consequence of the weak convergence of the local experiments to the Gaussian shift experiment.

The corresponding one-sided tests are easily derived along the same lines. Theorem 2 shows that our test has the same asymptotic behavior as the LR, W and RS tests. As already explained in the Introduction, our tests improve on these classical proposals by the non-necessity of estimating \( \nu \) (neither under the null nor under the alternative), the freedom of choice among root- \( n \) consistent estimators \( \hat{\mu}^{(n)} \) and \( \hat{\Sigma}^{(n)} \) and the ensuing simplicity. Yet another advantage of our Le Cam approach lies in the fact that Part (ii) of Theorem 2 makes it possible to easily write down the power of \( \phi^{(n)}_{\nu_0} \). Denoting by \( \Phi \) the cumulative distribution function of the standard Gaussian distribution, the asymptotic power of \( \phi^{(n)}_{\nu_0} \) under local alternatives of the form \( \bigcup_{\mu \in \mathbb{R}^k} \bigcup_{\Sigma \in S_k} P^{(n)}_{(\mu, \Sigma, \nu_0 + n^{-1/2} \tau_3^{(n)})} \) (\( \tau_3 := \lim_{n \to \infty} \tau_3^{(n)} \)) is then given by

\[
1 - \Phi \left( z_{\alpha/2} - \tau_3 \sqrt{\Gamma^*_{33}(\mu, \Sigma, \nu_0)} \right) + \Phi \left( -z_{\alpha/2} - \tau_3 \sqrt{\Gamma^*_{33}(\mu, \Sigma, \nu_0)} \right),
\]

and by

\[
1 - \Phi \left( z_{\alpha} - \tau_3 \sqrt{\Gamma^*_{33}(\mu, \Sigma, \nu_0)} \right) \quad \text{and} \quad \Phi \left( -z_{\alpha} - \tau_3 \sqrt{\Gamma^*_{33}(\mu, \Sigma, \nu_0)} \right)
\]

in the respective one-sided tests against \( \mathcal{H}_{\nu_0}^+ : \nu > \nu_0 \) and \( \mathcal{H}_{\nu_0}^- : \nu < \nu_0 \).

We conclude this section by briefly quantifying the loss in power due to the estimation of the scatter parameter. The ULAN result in Theorem 4 is about the “unspecified scatter” model, but it evidently entails ULAN for specified scatter. Thus, locally and asymptotically optimal tests for \( \mathcal{H}_0 : \nu = \nu_0 \) under specified scatter \( \Sigma \) reject \( \mathcal{H}_0 \) (at asymptotic level \( \alpha \)) whenever

\[ |Q^{(n)}_{\nu_0, \Sigma}| > z_{\alpha/2}, \]

with

\[ Q^{(n)}_{\nu_0, \Sigma} := \Delta_3^{(n)}(\hat{\mu}^{(n)}, \Sigma, \nu_0) \sqrt{\Gamma_{33}(\nu_0)}, \]

where \( \hat{\mu}^{(n)} \) is a sequence of estimators satisfying Assumption A. Along the same lines as in the proof of Theorem 4, one can show that the asymptotic behavior of \( Q^{(n)}_{\nu_0, \Sigma} \) under the local alternatives is \( \mathcal{N}(\tau_3 \sqrt{\Gamma_{33}(\nu_0)}, 1) \). The non-centrality parameters in the asymptotic non-null distributions of \( Q^{(n)}_{\nu_0} \) and \( Q^{(n)}_{\nu_0, \Sigma} \) allow for computing the efficiency loss due to an unspecified \( \Sigma \), which is simply the difference of those local shifts, hence

\[
\tau_3 \left( \sqrt{\Gamma_{33}(\nu_0)} - \sqrt{\Gamma^*_{33}(\mu, \Sigma, \nu_0)} \right).
\]

The positive definiteness of \( \Gamma_{22}(\mathcal{V}) \) in \( \Gamma^*_{33}(\mu, \Sigma, \nu_0) \) confirms the unsurprising fact that this loss is strictly positive. Quite remarkably, it does not depend on the scale \( \sigma := \sqrt{\Sigma} \) in the one-dimensional setup.

4. One-step estimation: locally and asymptotically optimal estimators for tail weight

In this section, we propose optimal one-step estimators for \( \nu \) using the ULAN structure and the efficient central sequence for tail weight. The main idea behind one-step estimation consists in adding to
an existing adequate preliminary estimator \( \hat{\nu}^{(n)} \) a quantity depending on a version of the efficient central sequence for \( \nu \). More precisely, Le Cam suggests estimating \( \nu \) by means of

\[
\hat{\nu}_{\text{Cam}}^{(n)} = \nu^{(n)} + n^{-1/2} \left( \Gamma^{*}_{33} \left( \hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, \hat{\nu}^{(n)} \right) \right)^{-1} \Delta_{3}^{(n)*} \left( \hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, \hat{\nu}^{(n)} \right), \tag{10}
\]

where \( \left( \hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, \hat{\nu}^{(n)} \right) \) is a preliminary estimator of \( (\mu, \Sigma, \nu) \) fulfilling Assumption A. Examples of such estimators for \( \mu \) and \( \Sigma \) have been given in Section 3.1, examples for \( \hat{\nu}^{(n)} \) will be discussed below. The following result states the asymptotic properties of the one-step estimator.

**Theorem 3.** Suppose that \( \hat{\mu}^{(n)} \) and \( \hat{\Sigma}^{(n)} \) satisfy Assumption A. Let \( \hat{\nu}_{\text{Cam}}^{(n)} \) be an estimator of \( \nu^{(n)} \) fulfilling also Assumption A and let \( \hat{\nu}_{\text{Cam}}^{(n)} \) be the one-step estimator given by (10). Then, under \( P^{(n)}_{\nu} \),

\[
n^{1/2}(\hat{\nu}_{\text{Cam}}^{(n)} - \nu) \overset{d}{\to} N \left( 0, (\Gamma^{*}_{33}(\nu))^{-1} \right)
\]

as \( n \to \infty \). Moreover, \( \hat{\nu}_{\text{Cam}}^{(n)} \) is the most efficient estimator for \( \nu \) under \( \bigcup_{\nu \in \mathbb{R}} \bigcup_{\nu \in \mathbb{R} \setminus \mathbb{N}} P^{(n)}_{\nu} \).

**Proof.** Let us start by showing that the asymptotic distribution under \( P^{(n)}_{\nu} \) of \( n^{1/2}(\hat{\nu}_{\text{Cam}}^{(n)} - \nu) \) is the same as that of \( (\Gamma^{*}_{33}(\nu))^{-1} \Delta_{3}^{(n)*}(\nu) \). From the asymptotic linearity of \( \Delta_{3}^{(n)*}(\nu) \) in Lemma 1 combined with Lemma 4.4 of Kreiss (1987) under Assumption A and from the continuity of \( \nu \mapsto \Gamma^{*}_{33}(\nu) \), we obtain that

\[
n^{1/2}(\hat{\nu}_{\text{Cam}}^{(n)} - \nu) = n^{1/2}(\hat{\nu}^{(n)} - \nu) + \left( \Gamma^{*}_{33} \left( \hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, \hat{\nu}^{(n)} \right) \right)^{-1} \Delta_{3}^{(n)*} \left( \hat{\mu}^{(n)}, \hat{\Sigma}^{(n)}, \hat{\nu}^{(n)} \right) + o_{P}(1)
\]

under \( P^{(n)}_{\nu} \) as \( n \to \infty \). The asymptotic behavior then directly follows thanks to the ULAN property. Efficiency of the estimator can be seen by noticing that the asymptotic variance coincides with the inverse of the efficient Fisher information.

Theorem 3 shows that, whatever the performance of the preliminary root-\( n \) consistent estimator \( \hat{\nu}^{(n)} \) and whatever the choice of root-\( n \) consistent estimators for location and scatter, the one-step estimator \( \hat{\nu}_{\text{Cam}}^{(n)} \) is as efficient as the maximum likelihood estimator (MLE). Thus we can make any root-\( n \) consistent estimator for tail weight as efficient as the MLE in a quite simple way. This is of course particularly interesting for estimators \( \hat{\nu}^{(n)} \) that have a closed-form expression or are easy to compute.

One of the most famous multivariate closed-form and root-\( n \) consistent estimators for tail weight goes back to Mardia (1970). In that seminal paper, Mardia has defined multivariate excess kurtosis as

\[
\kappa = \frac{E \left[ \left( (X - \mu)^T \Sigma^{-1} (X - \mu) \right)^2 \right]}{k(k+2)} - 1
\]

and its sample analogue as

\[
\hat{\kappa}^{(n)} = \frac{1}{k(k+2)n} \sum_{i=1}^{n} \left( (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}) \right)^2 - 1,
\]

where \( \bar{X} \) is the sample mean vector and \( S \) the sample covariance matrix. In the multivariate Student case, the theoretical value of \( \kappa \) exists if \( \nu > 4 \) and is given by \( \kappa = \frac{2}{\nu-4} \). Reversing the latter relation yields \( \nu = \frac{2}{\kappa} + 4 \), and so the kurtosis-based estimator for tail weight admits an exact expression, namely

\[
\hat{\nu}^{(n)}_{\text{Mardia}} = \frac{2}{\hat{\kappa}^{(n)}} + 4.
\]
In dimension one, the Mardia estimator of course coincides with the well-known moment estimator. Henze (1994) proved that \( \hat{\nu}_{Mardia} \) is a root-\( n \) consistent estimator for \( \nu > 8 \), entailing the root-\( n \) consistency of \( \hat{\nu} \). By the one-step estimation procedure, we can render Mardia’s estimator as efficient as the MLE in the multivariate Student case. The “advantage” of the restriction \( \nu > 8 \) is that we can use \( \hat{\mu} = \bar{X} \) and \( \hat{\Sigma} = S(1 - 2/\hat{\nu}) \), and so the new estimator, which we shall denote by \( \hat{\nu}_{MarCam} \), is clearly easier to calculate than the MLE. It thus represents an attractive alternative to the MLE for moderately heavy-tailed (\( \nu > 8 \)) distributions. In order to know if \( \hat{\nu}_{MarCam} \) can be used in a given situation, we suggest to first use our test \( \phi_{\nu=8} \) under one-sided form (see Section 6, where we apply this scheme on financial data).

In the one-dimensional setup, these moment conditions can be avoided by having recourse to the McElroy (2007) estimator \( \hat{\nu}_{McElroy} \), which is suited for small values of \( \nu \). In that paper, McElroy develops a root-\( n \) consistent estimator for tail weight by using log-moments. Defining \( \bar{X}_{\log} := n^{-1} \sum_{i=1}^{n} \log |X_i| \) and \( s_{\log} := n^{-1} \sum_{i=1}^{n} (\log |X_i| - \bar{X}_{\log})^2 \), the latter is given by

\[
\hat{\nu}_{McElroy} = g^{-1}(s_{\log})
\]

with

\[
g(x) = \frac{1}{4} \left( \psi \left( \frac{x}{2} \right) + \psi \left( \frac{1}{2} \right) \right),
\]

where \( \psi \) is the digamma function. There exists no exact expression for \( \psi^{-1} \), but \( g^{-1} \) can easily be computed numerically. Despite the non-exact expression we obtain, the resulting one-step estimator \( \hat{\nu}_{McECam} \), using the sample median and the MAD as location and scale estimators, is a good competitor for the MLE and is valid for any values of \( \nu \). Unfortunately, this estimator does not exist for higher dimensions, nor do the authors know of any closed-form estimator that is root-\( n \) consistent and avoids moment assumptions. However, would any such estimator be constructed for multivariate Student distributions, then our one-step estimation procedure would allow to render it fully efficient; we hope this might stimulate future research.

We conclude this section by briefly addressing the case of the most famous (one-dimensional) closed-form estimators for \( \nu \) available in the literature, namely the Hill estimator (Hill 1975) and the Pickands estimator (Pickands 1975). Both estimators, and their numerous variants, rely on the top \( k \) order statistics of the empirical distribution and satisfy (under some regularity conditions)

\[
\sqrt{k} \left( \hat{\nu}_{k} - \nu \right) = O_p(1),
\]

provided that \( n \to \infty \), \( k \to \infty \) and \( n/k \to \infty \). Since this rate is slower than the standard root-\( n \) rate, estimators of that type cannot be used in the present framework.

5. Simulation studies

In order to investigate the finite-sample properties of the inferential procedures proposed in this paper, we have conducted a Monte Carlo simulation study. The experimental designs and results for both our tests (under two-sided form) and our one-step estimators are given in Sections 5.1 and 5.2, respectively. The code for the simulation study has been written in R and is available from the authors upon request.

5.1. Finite-sample behavior of our new test based on the “Le Cam methodology”

By construction, our new tests \( \phi_{\nu} \) are as powerful as the LR, W and RS tests but improve on the latter classical tests by their simplicity and computational speed. In order to corroborate this assertion,
Table 1: Rejection frequencies (out of $N = 500$ replications) and total computation time in two distinct univariate situations for our test $\phi_{\nu_0}^{(n)}$, the likelihood ratio (LR) test, the Wald (W) test and the Rao score (RS) or Lagrange multiplier test. The nominal level is $\alpha = 0.05$.

| Test  | $\nu_0 = 1$ | $\nu = 1.1$ | $\nu = 1.2$ | $\nu = 1.3$ | Time  | $\nu_0 = 8$ | $\nu = 7$ | $\nu = 6$ | $\nu = 5$ | Time  |
|-------|-------------|-------------|-------------|-------------|-------|-------------|-------------|-------------|-------------|-------|
| LR    | 0.044       | 0.206       | 0.600       | 0.864       | 54 s  | 0.064       | 0.116       | 0.300       | 0.706       | 177 s |
| W     | 0.044       | 0.160       | 0.530       | 0.822       | 29 s  | 0.054       | 0.194       | 0.440       | 0.828       | 131 s |
| RS    | 0.044       | 0.180       | 0.564       | 0.830       | 42 s  | 0.060       | 0.140       | 0.350       | 0.750       | 85 s  |
| $\phi_{\nu_0}^{(n)}$ | 0.042       | 0.168       | 0.558       | 0.828       | 21 s  | 0.054       | 0.170       | 0.394       | 0.796       | 41 s  |

In all cases, the observations have location $\mu = 0$ and scale $\sigma = 1$, and the nominal level is $\alpha = 5\%$. The rejection frequencies as well as the total computation time on R are given in Table 1 (we have used the function `optim` in order to calculate the MLE on R). Inspection of this table clearly shows that our test is equivalent in terms of power to the other tests, but indeed outperforms them in terms of duration. This becomes all-the-more obvious in the second setup with $n = 1000$ observations. Since the complexity of ML estimation for the Student $t$ distribution increases with the dimension, the advantage of our not-ML procedure should become even further striking.

After this necessary comparison with the classical tests, we now come to a more detailed investigation of our tests. First we study the finite-sample effects of the choice of the nuisance parameter estimators $\hat{\mu}^{(n)}$ and $\hat{\Sigma}^{(n)}$ on the power of $\phi_{\nu_0}^{(n)}$. Of course, asymptotically every such pair of estimators performs equally well but, in finite-sample practice, there may well exist systematic differences (such finite-sample differences are also present in our previous comparison with the LR, W and RS tests, but the results do not indicate a domination of one particular test). Section 3.1 discusses examples of multivariate estimators for location and scatter, and we have compared three such couples: the sample mean $\bar{X}$ with the sample covariance matrix $S(1-2/\nu_0)$ (combination called “MeCov”), the Möttönen-Oja spatial median with the adjusted Tyler shape matrix (combination “MOT”), and the MCD estimators for location and scatter (combination “MCD”). Due to the presence of the sample covariance matrix which excludes low values of the tail parameter, we have tested the null $\nu_0 = 5$ against $\nu = 6, 7, 8$, for the two sample sizes $n = 500$ and $n = 1000$ and with location $\mu = 0$ and scatter $\Sigma = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$. We have generated $N = 2,500$ replications of each experiment, and the results are provided in Table 2. One clearly sees a ranking between the three couples of estimators: MeCov is the most powerful and MCD the weakest. This is not surprising, as MCD-based methods are known to exhibit very moderate power. Note however that the differences in performance decrease when passing from 500 to 1000 observations, which is consistent with the asymptotic equivalence of all estimators. Thus, for large sample sizes (as is usually the case with
Table 2: Rejection frequencies (out of $N = 2,500$ replications), under 2-variate Student $t$ densities with location $\mu = \mathbf{0}$ and scatter $\Sigma$ with $\Sigma_{11} = 5, \Sigma_{21} = \Sigma_{12} = 3$ and $\Sigma_{22} = 2$, for the null hypothesis $\nu_0 = 5$ against $\nu = 6, 7, 8$, of the optimal test $\hat{\nu}_5^{(n)}$ for various couples of estimators $(\hat{\mu}^{(n)}, \hat{\Sigma}^{(n)})$: sample mean and sample covariance matrix (MeCov), Möttönen-Oja spatial median and adjusted Tyler estimator of shape (MOT), and MCD-based estimators for location and scatter (MCD). The nominal level is $\alpha = 0.05$.

|       | $n = 500$ |       | $n = 1000$ |
|-------|-----------|-------|------------|
|       | $\nu = 5$ | $\nu = 6$ | $\nu = 7$ | $\nu = 8$ | $\nu = 5$ | $\nu = 6$ | $\nu = 7$ | $\nu = 8$ |
| MeCov | 0.0444    | 0.1748 | 0.4508    | 0.6860    | 0.0486    | 0.3276    | 0.7560    | 0.9472    |
| MOT   | 0.0504    | 0.0980 | 0.3048    | 0.5428    | 0.0464    | 0.2584    | 0.6824    | 0.9148    |
| MCD   | 0.0508    | 0.0856 | 0.2712    | 0.5000    | 0.0472    | 0.2228    | 0.6416    | 0.8960    |
Figure 1: Power curves (out of $N = 2,500$ replications), under 2-variate Student $t$ densities with location $\mu = 0$ and scatter $\Sigma$ with $\Sigma_{11} = 5/16, \Sigma_{21} = \Sigma_{12} = 3/8$ and $\Sigma_{22} = 17/16$, of the optimal test $\phi_{\nu_0}^{(n)}$ for the three null hypotheses $\nu_0 = 5, 7$ and $10$ against $\nu = \nu_0 \pm 1$, $\nu = \nu_0 \pm 2$ and $\nu = \nu_0 \pm 3$. The sample size is $n = 1000$ and the nominal level $\alpha = 0.05$.

| Size  | $\nu$ | $\hat{\nu}_{\text{McElroy}}^{(n)}$ | $\hat{\nu}_{\text{McECam}}^{(n)}$ | Size  | $\nu$ | $\hat{\nu}_{\text{Mardia}}^{(n)}$ | $\hat{\nu}_{\text{MarCam}}^{(n)}$ |
|-------|-------|----------------------------------|----------------------------------|-------|-------|----------------------------------|----------------------------------|
| $n = 200$ | 1     | 0.0422                           | 0.0186                           | $n = 2000$ | 9     | 3.1757                           | 1.3608                           |
| $n = 500$ | 1     | 0.0132                           | 0.0064                           | $n = 5000$ | 9     | 1.1856                           | 0.5347                           |
| $n = 1000$ | 1.5   | 0.0446                           | 0.0140                           | $n = 2000$ | 13    | 9.7261                           | 6.9017                           |
| $n = 2000$ | 2     | 0.0844                           | 0.0311                           | $n = 5000$ | 13    | 3.3589                           | 2.1437                           |

Table 3: Empirical MSE (out of $N = 2,500$ replications), under univariate Student $t$ densities with location $\mu = 0$ and scale $\sigma = 1$ (McElroy setting) and under 2-variate Student $t$ densities with location $\mu = 0$ and scatter $\Sigma$ with $\Sigma_{11} = 5, \Sigma_{21} = \Sigma_{12} = 3$ and $\Sigma_{22} = 2$ (Mardia setting), of the McElroy and Mardia estimators and of the corresponding one-step estimators for different sample sizes.
Figure 2: Power curves (out of $N = 2,500$ replications), under 2-variate Student $t$ densities with location $\mu = 0$ and scatter $\Sigma$ with $\Sigma_{11} = 5, \Sigma_{21} = \Sigma_{12} = 4$ and $\Sigma_{22} = 5$, of the optimal test $\phi_{\nu_0}^{(n)}$ for the null hypothesis $\nu_0 = 7$ against $\nu = 4, 5, 6, 8, 9$ and 10, for various sample sizes $n = 500$ (solid line), $n = 2000$ (dashed line), $n = 3000$ (dotted line) and $n = 5000$ (dash-dotted line). The nominal level is $\alpha = 0.05$.

Figure 3: Power curves (out of $N = 2,500$ replications), under $k$-variate Student $t$ densities with location $\mu = 0$ and scatter $\Sigma = I_k$, of the optimal test $\phi_{\nu_0}^{(n)}$ for the null hypothesis $\nu_0 = 5$ against $\nu = 2, 3, 4, 6, 7$ and 8, for various dimensions $k = 1$ (solid line), $k = 2$ (dashed line) and $k = 5$ (dotted line). The sample size is $n = 1000$ and the nominal level $\alpha = 0.05$. 
Lemma 2. and thus completes the proof of Theorem 1.

East (AEX, ATX, FTSE, DAX, CAC40, SMI, MIB and TA100), and East Asia and Oceania (HgSg, Nikkei, StrTim, SSEC, BSE, KLSE, KOSPI and AllOrd). The sample consists of 2536 observations, from January 4, 2000 to September 22, 2009. The same data has already been analyzed in Dominicy and Veredas (2013) and Dominicy et al. (2013). We refer the reader to Table 5 of Dominicy and Veredas (2013) for some information about the descriptive statistics of each return series, and to Dominicy et al. (2013) for a detailed description of the filtering used and the data set’s criticisms. The purpose of our study here is to estimate the tail index under the Student $t$ assumption using the Student-adaptive methods developed in the present paper.

First, we consider the right-sided test for tail weight $\phi_{\nu_0=8}^{(\nu)}$ (with the most efficient and simplest MeCov combination for estimating the location and scatter nuisance parameters) and obtain a $p$-value equal to $6.8 \times 10^{-7}$. Given the strong rejection of $H_0 : \nu_0 = 8$ against $H_1^\nu : \nu > 8$ at any possible level, we can apply the improved one-step Mardia estimator $\nu_{\text{MarCam}}$ (using as well the MeCov combination) and find a value of 9.75, which clearly differs from the $\nu_{\text{Mardia}} = 8.72$ provided by the classical Mardia estimator. While 8.72 is dramatically rejected by our test, with a $p$-value of 0.0005, the value of 9.75 is neither rejected by a right-sided nor by a left-sided test ($p$-values 0.17 and 0.83, respectively), hence fits the data set well. In order to refine our analysis of the data set at hand, we have constructed via our right- and left-sided tests (asymptotic) confidence intervals for the tail weight parameter, resulting in [9.44, 10.96] at the 95% confidence level and [9.13, 11.29] at the 99% confidence level. Both intervals contain our value of 9.75 and confirm the value of 9.51 derived in Dominicy et al. (2013).

7. Appendix

The following lemma establishes the quadratic mean differentiability of $(\mu', (\text{vech}(\Sigma))^\prime)^\nu \rightarrow f_{\mu,\Sigma,\nu}^{1/2}(x)$ and thus completes the proof of Theorem 1.

**Lemma 2.** For any $\mu \in \mathbb{R}^k$, $\Sigma \in S_k$ and $\nu > 0$, define

$$D_{\mu,\nu} f_{\mu,\nu}^{1/2}(x) = f_{\mu,\nu}^{1/2}(x) \left(1 + k/\nu \right) \frac{1}{2 \left(1 + \|\Sigma^{-1/2}(x - \mu)\|^2/\nu \right)} \Sigma^{-1}(x - \mu),$$

$$D_{\text{vech}(\Sigma),\nu} f_{\mu,\nu}^{1/2}(x) = \frac{1}{4} f_{\mu,\nu}^{1/2}(x) P_k(\Sigma^{\otimes 2})^{-1/2} \left(1 + k/\nu \right) \Sigma^{-1/2}(x - \mu) \Sigma^{-1/2} - I_k,$$

and

$$D_{\nu} f_{\mu,\nu}^{1/2}(x) = f_{\mu,\nu}^{1/2}(x) \left(\frac{\nu k/\nu}{2\nu k} - 1 + \log \left(1 + \|\Sigma^{-1/2}(x - \mu)\|^2/\nu \right) + \nu + k \right) \frac{1}{4 \nu^2} \frac{\|\Sigma^{-1/2}(x - \mu)\|^2}{1 + \|\Sigma^{-1/2}(x - \mu)\|^2/\nu}.$$

Then, as $(s', (\text{vech}(H))^\prime, t) \rightarrow (0', 0', 0)$ such that $\Sigma + H \in S_k$ and $\nu + t > 0$,

(i) \( \int_{\mathbb{R}^k} \left\{ f_{\mu+s,\nu}^{1/2}(x) - f_{\mu,\nu}^{1/2}(x) - s' D_{\mu,\nu} f_{\mu,\nu}^{1/2}(x) \right\}^2 \text{d}x = o(||s||^2) \),

(ii) \( \int_{\mathbb{R}^k} \left\{ f_{\mu+s,\nu}^{1/2}(x) - f_{\mu,\nu}^{1/2}(x) - (\text{vech}(H))^\prime D_{\text{vech}(\Sigma),\nu} f_{\mu+s,\nu}^{1/2}(x) \right\}^2 \text{d}x = o(||\text{vech}(H)||^2) \),

(iii) \( \int_{\mathbb{R}^k} \left\{ f_{\mu+s,\nu}^{1/2}(x) - f_{\mu+s,\nu}^{1/2}(x) - t D_{\nu} f_{\mu+s,\nu}^{1/2}(x) \right\}^2 \text{d}x = o(t^2) \),
\( (iv) \int_{\mathbb{R}^k} \left\{ D_{\nu}^{1/2} \mu, \Sigma + Hu, \nu (x) - D_{\nu}^{1/2} \mu, \Sigma, \nu (x) \right\}^2 dx = o(1), \)

\( (v) \int_{\mathbb{R}^k} \left\| D_{\nu}^{1/2} \mu, \Sigma + Hu, \nu (x) - D_{\nu}^{1/2} \mu, \Sigma, \nu (x) \right\|^2 dx = o(1), \)

\( (vi) \int_{\mathbb{R}^k} \left\{ f_{\mu, \Sigma + Hu, \nu + t} (x) - f_{\mu, \Sigma, \nu} (x) \left( \frac{s}{v} \right)^{vech (H)} \left( D_{\nu}^{1/2} \mu, \Sigma + Hu, \nu (x) \right) \right\}^2 dx = o \left( \left\| \left( \frac{s}{v} \right)^{vech (H)} \right\|^2 \right). \)

**Proof.** In this proof, all \( o(\| \cdot \|) \) and \( O(\| \cdot \|) \) quantities are taken as \( \| \cdot \| \to 0 \). For Points (i) and (ii), we refer to Lemma A.1 in Hallin and Paindaveine (2006); the other points will be established in what follows. 

(iii) We can assume w.l.o.g. that \( t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right] \). Indeed, \( \nu + t \) then belongs to \( \mathbb{R}^+ \), as \( \nu \) varies in \( \mathbb{R}^+ \). Letting \( y := (\Sigma + H)^{-1/2} (x - \mu - s) \), the left-hand side in (iii) can be rewritten as

\[
\int_{\mathbb{R}^k} \left\{ \sqrt{c_{\nu + t, k}} \left( 1 + \frac{\|y\|^2}{\nu + t} \right)^{\frac{\nu + k}{4}} - \sqrt{c_{\nu, k}} \left( 1 + \frac{\|y\|^2}{\nu} \right)^{\frac{\nu + k}{4}} \times \left( 1 + t \left( \frac{c_{\nu, k}}{2c_{\nu, k}} - \frac{1}{4} \log \left( 1 + \frac{\|y\|^2}{\nu} \right) + \frac{\nu + k}{4\nu^2} \frac{\|y\|^2}{1 + \frac{\|y\|^2}{\nu}} \right) \right) \right\}^2 dy.
\]

(11)

It follows from the triangular inequality that the integral can be bounded by \( C(T_1 + T_2 + T_3) \), where \( C \) is some positive constant,

\[
T_1 := \left( \sqrt{c_{\nu + t, k}} - \sqrt{c_{\nu, k}} - t \frac{c_{\nu, k}}{2\sqrt{c_{\nu, k}}} \right)^2 \int_{\mathbb{R}^k} \left( 1 + \frac{\|y\|^2}{\nu + t} \right)^{\frac{\nu + k}{4}} - \left( 1 + \frac{\|y\|^2}{\nu} \right)^{\frac{\nu + k}{4}} dy,
\]

\[
T_2 := t^2 T_{2a} := t^2 \left( \frac{c_{\nu, k}}{2\sqrt{c_{\nu, k}}} \right)^2 \int_{\mathbb{R}^k} \left\{ \left( 1 + \frac{\|y\|^2}{\nu + t} \right)^{\frac{\nu + k}{4}} - \left( 1 + \frac{\|y\|^2}{\nu} \right)^{\frac{\nu + k}{4}} \right\}^2 dy,
\]

and

\[
T_3 := c_{\nu, k} \int_{\mathbb{R}^k} \left\{ \left( 1 + \frac{\|y\|^2}{\nu + t} \right)^{\frac{\nu + k}{4}} - \left( 1 + \frac{\|y\|^2}{\nu} \right)^{\frac{\nu + k}{4}} \right\} - t \left( 1 + \frac{\|y\|^2}{\nu} \right)^{\frac{\nu + k}{4}} \times \left( -\frac{1}{4} \log \left( 1 + \frac{\|y\|^2}{\nu} \right) + \frac{\nu + k}{4\nu^2} \frac{\|y\|^2}{1 + \frac{\|y\|^2}{\nu}} \right) \right\}^2 dy.
\]

Clearly, routine Taylor series arguments directly yield

\[
T_1 = \frac{1}{c_{\nu + t, k}} \left( \sqrt{c_{\nu + t, k}} - \sqrt{c_{\nu, k}} - t \frac{c_{\nu, k}}{2\sqrt{c_{\nu, k}}} \right)^2 = o(t^2).
\]

Now, working in spherical coordinates \((r, u) = (\|y\|, \frac{y}{\|y\|})\) leads to

\[
T_{2a} = \omega_k \left( \frac{c_{\nu, k}}{2\sqrt{c_{\nu, k}}} \right)^2 \int_0^\infty r^{k-1} \left\{ \left( 1 + \frac{r^2}{\nu + t} \right)^{\frac{\nu + k}{4}} - \left( 1 + \frac{r^2}{\nu} \right)^{\frac{\nu + k}{4}} \right\}^2 dr,
\]

\[20\]
where $\omega_k$ is the surface area of the $k$-dimensional unit hypersphere. We have to prove that $T_{2\alpha}$ is $o(1)$, as $t$ goes to zero. To do so, it is sufficient to prove

**Lemma 2A.** For all $t \in \left[\frac{-\nu}{2}, \frac{\nu}{2}\right]$, set

$$h_t(r) = r^{k-1} \left(1 + \frac{r^2}{\nu + t}\right)^{\frac{-\nu+k+1}{2}}, \quad r \geq 0.$$  

There exists an integrable function $g : \mathbb{R}^+ \to \mathbb{R}^+$ independent of $t$ such that for all $t \in \left[\frac{-\nu}{2}, \frac{\nu}{2}\right]$

$$h_t(r) \leq g(r) \quad \forall r \geq 0.$$  

**Proof of Lemma 2A.** For sufficiently large $r$, we will establish the monotone decreasing behavior of the mapping $t \mapsto h_t(r)$. Basic computations yield

$$\frac{\partial}{\partial t} h_t(r) = \frac{1}{2} h_t(r) m_t(r) := \frac{1}{2} h_t(r) \left[-\log \left(1 + \frac{r^2}{\nu + t}\right) + \frac{\nu + t + k}{\nu + t} \left(1 + \frac{r^2}{\nu + t}\right)^{-1}\right].$$

Since $t \in \left[\frac{-\nu}{2}, \frac{\nu}{2}\right]$ by assumption, we easily see that

$$\frac{2}{3\nu} \leq \frac{1}{\nu + t} \leq \frac{2}{\nu}. \quad (12)$$

In view of the inequalities in (12) and the monotonicity property of $x \mapsto -\log(x)$ over $\mathbb{R}^+$, we have the following bounds

$$-\log \left(1 + \frac{r^2}{\nu + t}\right) \leq -\log \left(1 + \frac{2r^2}{3\nu}\right) \quad \text{and} \quad \frac{\nu + t + k}{\nu + t} \leq 1 + \frac{2k}{\nu}.$$  

Combining the latter bounds with the fact that $\frac{x}{1+x} \leq 1$ for all $x \geq 0$, we obtain

$$m_t(r) \leq -\log \left(1 + \frac{r^2}{\nu + t}\right) + 1 + \frac{k}{\nu + t} \leq -\log \left(1 + \frac{2r^2}{3\nu}\right) + 1 + \frac{2k}{\nu}.$$  

Now, set $M := \sqrt{\frac{3\nu}{2} \left(\exp(1 + \frac{2k}{\nu}) - 1\right)}$. The monotone decreasing nature of $x \mapsto -\log(x)$ then yields that, for all $r \geq M$,

$$m_t(r) \leq -\log \left(1 + \frac{2r^2}{3\nu}\right) + 1 + \frac{2k}{\nu} \leq -\log \left(1 + \frac{2M^2}{3\nu}\right) + 1 + \frac{2k}{\nu} = 0 \quad \forall t \in \left[-\frac{\nu}{2}, \frac{\nu}{2}\right],$$

implying that for all $r \geq M$,

$$\frac{\partial}{\partial t} h_t(r) \leq 0 \quad \forall t \in \left[-\frac{\nu}{2}, \frac{\nu}{2}\right].$$  

It follows that for all $r \geq M$, $h_t(r)$ is a monotonically decreasing function in $t$ and thus

$$h_t(r) \leq h_{t(-\frac{\nu}{2})}(r) \quad \forall r \geq M, \quad \forall t \in \left[-\frac{\nu}{2}, \frac{\nu}{2}\right].$$  

By splitting $\mathbb{R}^+$ into $[0, M]$ and $[M, \infty)$, $h_t(r)$ takes on the guise

$$h_t(r) = h_t(r)I_{[0, M]}(r) + h_t(r)I_{[M, \infty)}(r) \leq r^{k-1}I_{[0, M]}(r) + h_{t(-\frac{\nu}{2})}I_{[M, \infty)}(r),$$

21
where the inequality follows from the monotonicity properties established above, combined with the fact that \( h_t(r) \leq r^{k-1} \) for all admissible \( r \) and \( t \). We conclude that by construction the dominating function \( g \) is defined as
\[
g(r) := r^{k-1} \left( I_{[0,M]}(r) + \left( 1 + \frac{r^2}{\nu^2} \right)^{-\frac{\nu+k}{4}} I_{[M,\infty]}(r) \right).
\]
The integrability of \( g \) follows from the Riemann criterion for improper integrals, combined with the assumption \( \nu > 0 \).

We now turn back to the proof of \( T_{2a} = o(1) \), as \( t \to 0 \). Indeed, using the triangular inequality and Lemma 2A, the integrand in \( T_{2a} \) is bounded by
\[
2h_t(r) + 2r^{k-1} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+k}{4}} \leq 2g(r) + 2r^{k-1} \left( 1 + \frac{r^2}{\nu} \right)^{-\frac{\nu+k}{4}},
\]
where the last function is independent of \( t \) and integrable on \( \mathbb{R}^+ \). Therefore it follows from Lebesgue’s dominated convergence theorem that \( T_{2a} = o(1) \), as \( t \) tends to zero and consequently \( T_2 \) is \( o(t^2) \).

To finish the proof of the quadratic mean differentiability of \( \nu \mapsto t^{1/2} \mu_{\Sigma,v}(\mathbf{x}) \), it remains to show that \( T_3 = o(t^2) \). To do so, define the function \( g_{\gamma} : (0,\infty) \to \mathbb{R}^+ \) by \( g_{\gamma}(\nu) = \left( 1 + \frac{\nu^2}{\nu + \gamma} \right)^{-\frac{\nu+k}{4}} \). By applying the mean value theorem, we can rewrite \( T_3 \) under the form
\[
T_3 = t^2 c_{\nu,k} T_{3a} := t^2 c_{\nu,k} \int_{\mathbb{R}^k} \left\{ g_{\gamma}^{\prime}(\gamma)\left|_{\gamma=\nu^t} - \gamma^t\right. - g_{\gamma}^{\prime}(\gamma)\left|_{\gamma=\nu^t} \right. \right\} d\gamma,
\]
where \( |\delta| < t \). Changing the Cartesian coordinates into spherical coordinates, \( T_{3a} \) takes on the guise
\[
T_{3a} = \omega_k \int_0^\infty r^{k-1} \left\{ \frac{1}{4} \left( 1 + \frac{r^2}{\nu^t + \delta} \right)^{-\frac{\nu+k}{4}} \left( -\log \left( 1 + \frac{r^2}{\nu^t + \delta} \right) + \frac{\nu + \delta + k}{\nu + \delta} \frac{r^2}{\nu + \delta} \left( 1 + \frac{r^2}{\nu + \delta} \right)^{-1} \right) \right\}^2 dr.
\]
As for \( T_{2a} \), we have to prove that \( T_{3a} \) is \( o(1) \), as \( t \) goes to zero. This will follow once more from the Lebesgue dominated convergence theorem, combined with the following lemma.

**Lemma 2B.** For all \( t \in \left[-\frac{\nu}{2},\frac{\nu}{2}\right] \), set
\[
q_t(r) = r^{k-1} \left( 1 + \frac{r^2}{\nu + t} \right)^{-\frac{\nu+k}{4}} \left( -\log \left( 1 + \frac{r^2}{\nu + t} \right) + \frac{\nu + t + k}{\nu + t} \frac{r^2}{\nu + t} \left( 1 + \frac{r^2}{\nu + t} \right)^{-1} \right)^2, \quad r \geq 0.
\]
There exists an integrable function \( j : \mathbb{R}^+ \to \mathbb{R}^+ \) independent of \( t \) such that for all \( t \in \left[-\frac{\nu}{2},\frac{\nu}{2}\right] \)
\[
q_t(r) \leq j(r) \quad \forall r \geq 0.
\]

**Proof of Lemma 2B.** Similar to the proof of Lemma 2A, we start by studying the monotonic behavior of the mapping \( t \mapsto q_t(r) \). Basic calculus manipulations lead to
\[
\frac{\partial}{\partial t} q_t(r) = 2 \frac{\partial}{\partial t} h_t(r) \left( \frac{1}{2} q_{1,t}(r) + q_{2,t}(r) \right),
\]
where \( h_t(t) \) is the function defined in Lemma 2A,
\[
q_{1,t}(r) = -\log \left( 1 + \frac{r^2}{\nu + t} \right) + \frac{\nu + t + k}{\nu + t} \frac{r^2}{\nu + t} \left( 1 + \frac{r^2}{\nu + t} \right)^{-1},
\]

22
Next, recall that by construction we have
\[
\frac{2}{3\nu} \leq \frac{1}{\nu + t} \leq \frac{2}{\nu},
\]
see [12]. The inequalities in [12], combined with the fact that \( \frac{x}{1+x} \leq 1 \) for all \( x \geq 0 \), entail that for all \( r \geq 0 \) and for all \( t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right] \)
\[
q_{2,t}(r) \geq -\frac{4k}{(\nu + t)^2} \geq -\frac{16k}{\nu^2} := -C. \tag{13}
\]

Now, performing similar manipulations as in Lemma 2A yields
\[
q_{1,t}(r) \leq -\sqrt{2C} \ ( < 0) \quad \forall r \geq N, \quad \forall t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right],
\]
where
\[
N := \sqrt{\frac{3\nu}{2}} \left( \exp \left( 1 + \frac{2k}{\nu} + \sqrt{2C} \right) - 1 \right).
\]

Taking into account that \( x \mapsto x^2 \) is a decreasing function over \( \mathbb{R}^- \) leads to
\[
\frac{\partial}{\partial t} h_t(r) \leq 0 \quad \forall r \geq M, \quad \forall t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right], \tag{15}
\]
with \( M := \sqrt{\frac{3\nu}{2}} \left( \exp \left( 1 + \frac{2k}{\nu} \right) - 1 \right) \). Since \( N \geq M \), the inequality in (15) remains in particular true for all \( t \) and all \( r \geq N \). In view of all these considerations, we deduce that
\[
\frac{\partial}{\partial t} q_t(r) \leq 0 \quad \forall r \geq N, \quad \forall t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right],
\]
implicating that
\[
q_t(r) \leq q_t(-\frac{\nu}{2})(r) \quad \forall r \geq N, \quad \forall t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right].
\]

Now, split \( \mathbb{R}^+ \) into \([0,N]\) and \([N,\infty)\) and partition the function \( q_t(r) \) in the same way. For all \( t \in \left[ -\frac{\nu}{2}, \frac{\nu}{2} \right] \), we can write
\[
q_t(r) = q_t(r) I_{[0,N]}(r) + q_t(r) I_{[N,\infty)}(r) \\
\leq q_t(r) I_{[0,N]}(r) + q_t(-\frac{\nu}{2}) I_{[N,\infty)}(r),
\]
where the last inequality follows from the monotonicity properties established above. Using the triangular inequality and the inequalities in [12], we can bound \( q_t(r) \) on the interval \([0,N]\) by
\[
q_t(r) \leq 2r^{k-1} \left\{ \log^2 \left( 1 + \frac{\nu^2}{\nu + t} \right) + \left( 1 + \frac{k}{\nu + t} \right)^2 \right\} \\
\leq 2r^{k-1} \left\{ \log^2 \left( 1 + \frac{2\nu^2}{\nu} \right) + \left( 1 + \frac{2k}{\nu} \right)^2 \right\}.
\]
Finally, for all \( r \geq 0 \), set
\[
j(r) = 2r^{k-1} \left\{ \log^2 \left( 1 + \frac{2r^2}{\nu} \right) + \left( 1 + \frac{2k}{\nu} \right)^2 \right\} I_{0,N}(r) + q(-\frac{r}{2})I_{N,\infty}(r).
\]

By construction, the function \( j \) is independent of \( t \) and dominates \( q_t \) on \( \mathbb{R}^+ \). To establish the integrability of \( j \), it is sufficient to show that \( q(-\frac{r}{2}) \) is integrable on \( \mathbb{R}^+ \); the integral of the first term of \( j \) is finite since integrated over a (compact) finite support. For all \( \nu > 0 \), we have by definition
\[
\int_0^\infty q(-\frac{r}{2}) \, dr = \int_0^\infty r^{k-1} \left( 1 + \frac{r^2}{\nu/2} \right)^{-\nu/(2k+2)} \left( -\log \left( 1 + \frac{r^2}{\nu/2} \right) + \frac{\nu/2 + k}{\nu/2} \frac{r^2}{\nu/2} \left( 1 + \frac{r^2}{\nu/2} \right)^{-1} \right)^2 \, dr.
\]
The triangular inequality allows us to bound the latter integral by \( 2(A_1 + A_2) \), with
\[
A_1 = \left( \frac{\nu/2 + k}{\nu/2} \right)^2 \int_0^\infty r^{k-1} \left( 1 + \frac{r^2}{\nu/2} \right)^{-\nu/(2k+2)} \left( \frac{r^2}{\nu/2} \right)^2 \left( 1 + \frac{r^2}{\nu/2} \right)^{-2} \, dr,
\]
and
\[
A_2 = \int_0^\infty r^{k-1} \left( 1 + \frac{r^2}{\nu/2} \right)^{-\nu/(2k+2)} \log^2 \left( 1 + \frac{r^2}{\nu/2} \right) \, dr = \left( \frac{\nu/2}{k} \right)^{k/2} \int_0^\infty r^{k-1} \left( 1 + r^2 \right)^{-\nu/(2k+2)} \log^2 (1 + r^2) \, dr,
\]
where the last equality is obtained by the change of variables \( r \to \frac{r}{\sqrt{\nu/2}} \). We easily see that
\[
A_1 \leq \left( \frac{\nu/2 + k}{\nu} \right)^2 \int_0^\infty r^{k-1} \left( 1 + \frac{r^2}{\nu/2} \right)^{-\nu/(2k+2)} \, dr = \left( \frac{\nu/2 + k}{\nu} \right)^2 \frac{1}{\omega_k c_{\nu/2,k}} < \infty.
\]
The rest of the proof relies on the fact that for any \( p > 0 \) there exists a constant \( M(p) > 0 \) such that
\[
\log^2 (1 + r^2) \leq (1 + r^2)^p \quad \forall r \geq M(p).
\] (16)

Splitting the integral into two integrals, one over \((0, M(p))\) and one over \([M(p), \infty)\), we can bound \( A_2 \) by
\[
\left( \frac{\nu}{2} \right)^{k/2} \int_0^{M(p)} r^{k-1} \left( 1 + r^2 \right)^{-\nu/(2k+2)} \log^2 (1 + r^2) \, dr + \left( \frac{\nu}{2} \right)^{k/2} \int_{M(p)}^\infty r^{k-1} \left( 1 + r^2 \right)^{-\nu/(2k+2)} \, dr,
\]
a quantity that is finite in view of the Riemann criterion for improper integrals, under the condition \( \nu > 4p \). Since \( p \) is arbitrary, we conclude that the dominating function \( j \) is integrable for all \( \nu > 0 \). \( \square \)

Using the preceding auxiliary lemma, we deduce that \( T_{3a} = o(1) \), as \( t \to 0 \). Indeed, a dominating (integrable, see the steps above for the second term) function of the integrand in \( T_{3a} \) is given by
\[
\frac{1}{8} j(r) + \frac{1}{4} r^{k-1} \left( 1 + \frac{r^2}{\nu} \right)^{-\nu/(2k+2)} \left( \log^2 \left( 1 + \frac{r^2}{\nu} \right) + \left( \frac{\nu + k}{\nu} \right)^2 \right).
\]
The Lebesgue dominating convergence theorem then allows us to interchange the limit and the integral. Consequently, \( T_{3a} \) is \( o(1) \). The claim in (iii) follows.

(iv) The triangular inequality allows us to bound the left-hand side of (iv) by \( 2(B_1 + B_2) \), where
\[
B_1 = \int_{\mathbb{R}^k} \left\{ D_{\nu} J^{1/2}_{\mu, \Sigma + H, \nu}(x) - D_{\nu} J^{1/2}_{\mu, \Sigma + H, \nu}(x) \right\}^2 \, dx
\]
and
\[ B_2 = \int_{\mathbb{R}^k} \left( D_\nu f_{\mu, \Sigma + H, \mu}(x) - D_\nu f_{\mu, \Sigma, \nu}(x) \right)^2 dx. \]

We can rewrite \( B_1 \) under the form
\[ B_1 = \int_{\mathbb{R}^k} \left( D_\nu f^1_{0, I_k, \nu}(x) - (\Sigma + H)^{-1/2} s \right) - D_\nu f^1_{0, I_k, \nu}(x) \right)^2 dx. \]

The quadratic mean continuity, combined with the fact that \( x \to D_\nu f^1_{0, I_k, \nu}(x) \) is square integrable over \( \mathbb{R}^k \) entails that \( B_1 = o(1) \), as \( s \) goes to zero. To show that \( B_2 = o(1) \), define the function \( g : \mathbb{R} \to \mathbb{R} \) by
\[ g(x) = \left( 1 + \frac{2^k}{x^2} \right)^{-\frac{1}{x^2} + \frac{1}{x^2}} \left( \frac{c_{\nu, k} - \frac{1}{2} \log(1 + x^2/\nu) + \frac{\nu k}{12} \frac{x^2}{1 + x^2} \right). \]

Letting \( y := \Sigma^{-1/2}(x - \mu) \), \( B_2 \) can be rewritten as
\[ B_2 = c_{\nu, k} \int_{\mathbb{R}^k} \left[ \frac{1}{|I_k + H\Sigma|^{1/4}} \right] \left| y \right|_{I_k + H\Sigma} - g(|y|) \right)^2 dy \]
\[ \leq 2 c_{\nu, k} (B_{2a} + B_{2b}), \]
where \( H\Sigma = \Sigma^{-1/2}H\Sigma^{-1/2} \),
\[ B_{2a} := \int_{\mathbb{R}^k} \left( \frac{1}{|I_k + H\Sigma|^{1/4}} - 1 \right)^2 g^2(|y|)_{I_k + H\Sigma} dy. \]
and
\[ B_{2b} := \int_{\mathbb{R}^k} \{g(|y|)_{I_k + H\Sigma} - g(|y|)\}^2 dy. \]

Putting \( z := (I_k + H\Sigma)^{-1/2}y \), \( B_{2a} \) takes on the guise
\[ \left( \frac{1}{|I_k + H\Sigma|^{1/4}} - 1 \right)^2 |I_k + H\Sigma|^{1/2} \int_{\mathbb{R}^k} g^2(|z|) dz = \frac{1}{c_{\nu, k}} \Gamma_{33}(\nu) \left| I_k + H\Sigma \right|^{1/4} - 1)^2. \]

The last equality is obtained by using
\[ \int_{\mathbb{R}^k} g^2(|z|) dz = \frac{1}{c_{\nu, k}} \int_{\mathbb{R}^k} f_{0, I_k, \nu}(z) (D_\nu(\log(f_{0, I_k, \nu}(z)))^2 dz = \frac{1}{c_{\nu, k}} \Gamma_{33}(\nu). \]

Since \(|A + B|^a = |A| + a |A| a \tr(A^{-1}B) + o(||B||) \) for all \( a \), we easily deduce that \( B_{2a} = o(1) \), as \( \text{vech}(H) \) tends to zero.

Now, working in spherical coordinates \((r, u) := (|y|, \frac{y}{|y|}) \) yields
\[ B_{2b} = \int_{S^{k-1}} \int_0^\infty \{g(r||u||_{I_k + H\Sigma}) - g(r)\}^2 r^{k-1} dr d\sigma(u) \]
\[ = \int_{S^{k-1}} \int_0^\infty \{g_{\exp} (\log(r) + \log (||u||_{I_k + H\Sigma})) - g_{\exp} (\log(r))\}^2 r^{k-1} dr d\sigma(u), \]
where \( g_{\exp}(x) := g(\exp(x)) \). Substituting \( z \) for \( \log r \) leads to
\[ B_{2b} = \int_{S^{k-1}} \int_{-\infty}^\infty \{g_{\exp} (z + \log (||u||_{I_k + H\Sigma})) - g_{\exp}(z)\}^2 \exp(kz) dz dr d\sigma(u). \]
Next, observe that \( g_{\exp} \in L^2(\mathbb{R}, \eta_k) \), where \( L^2(\mathbb{R}, \eta_k) \) is the space of square integrable functions with respect to the Lebesgue measure with weight \( \exp(kx) \) on \( \mathbb{R} \). By using the fact that \( \log(\|u\|_{L_k + H}) = O(\|\text{vech}(H)\|) \) for all \( u \), the quadratic mean continuity implies that

\[
\int_{-\infty}^{\infty} \left( g_{\exp}(z + \log(\|u\|_{L_k + H})) - g_{\exp}(z) \right)^2 \exp(kz) \, dz = o(1)
\]

for all \( u \in S^{k-1} \). Therefore, it follows from Lebesgue’s dominated convergence theorem that \( B_{2b} = o(1) \). The claim in \((iv)\) follows.

\((v)\) As \( x \to D_{\text{vech}(\Sigma)} f_{\mu, \Sigma, \nu}^{1/2}(x) \) is square integrable over \( \mathbb{R}^k \), this assertion is an immediate consequence of the quadratic mean continuity.

\((vi)\) To prove \((vi)\), observe that the left-hand side is bounded by \( C(S_1 + S_2 + S_3 + t^2 S_4 + \|\text{vech}(H)\|^2 S_5) \), where \( C \) is some positive constant,

\[
S_1 := \int_{\mathbb{R}^k} \left\{ \int_{\mu + s \Sigma + H, \nu}^{1/2}(x) - \int_{\mu + s \Sigma + H, \nu}^{1/2}(x) - s \int_{\mu + s \Sigma + H, \nu}^{1/2}(x) \right\}^2 \, dx,
\]

\[
S_2 := \int_{\mathbb{R}^k} \left\{ \int_{\mu + s \Sigma}^{1/2}(x) - \int_{\mu + s \Sigma}^{1/2}(x) - \int_{\mu + s \Sigma + H, \nu}^{1/2}(x) \right\}^2 \, dx,
\]

\[
S_3 := \int_{\mathbb{R}^k} \left\{ \int_{\mu + s \Sigma + t \Sigma + H, \nu}^{1/2}(x) - \int_{\mu + s \Sigma + t \Sigma + H, \nu}^{1/2}(x) \right\}^2 \, dx,
\]

\[
S_4 := \int_{\mathbb{R}^k} \left\{ \int_{\mu + s \Sigma + H, \nu + t}^{1/2}(x) - \int_{\mu + s \Sigma + H, \nu + t}^{1/2}(x) \right\}^2 \, dx,
\]

and

\[
S_5 := \int_{\mathbb{R}^k} \left\{ \int_{\mu + s \Sigma + H, \nu + t}^{1/2}(x) - \int_{\mu + s \Sigma + H, \nu + t}^{1/2}(x) \right\}^2 \, dx.
\]

The result then follows from \((i)\), \((ii)\), \((iii)\), \((iv)\) and \((v)\). \(\square\)

Acknowledgments

The research of Christophe Ley, who is also a member of ECARES, is supported by a Mandat de Chargé de Recherche from the Fonds National de la Recherche Scientifique, Communauté française de Belgique, whereas that of Anouk Neven is supported by the Fonds National de la Recherche, Luxembourg (Project Reference 4086487). Both authors would like to thank Yves Dominicy and David Veredas for providing them with the financial data used in Section 6 and Giovanni Peccati and Enrique Sentana for interesting comments and discussions.

References

References

[1] Blattberg, R.C., and Gonedes, N.J. (1974). A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *J. Bus. Univ. Chicago* 47, 244-280.

[2] Boothe, P., and Glassman, D. (1987). The statistical distribution of exchange rates. *J. Int. Econ.* 22, 297–319.

[3] Box, J.F. (1981). Gosset, Fisher, and the t Distribution. *Amer. Statist.* 35, 61–66.
[4] Cator, E.A. and Lopuhaä, H.P. (2012). Central limit theorem and influence function for the MCD estimators at general multivariate distributions. Bernoulli 18, 520–551.

[5] Chien, J.T. (2002). A Bayesian prediction approach to robust speech recognition and online environmental testing. West. J. Speech Commun. 37, 321-334.

[6] Cornish, E.A. (1954). The multivariate $t$-distribution associated with a set of normal sample deviates. Aust. J. Phys. 7, 531-542.

[7] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. Ser. B 39, 1–38.

[8] Dominicy, Y., and Veredas, D. (2013). The method of simulated quantiles. J. Econometrics 172, 235-247.

[9] Dominicy, Y., Ogata, H., and Veredas, D. (2013). Inference for vast dimensional elliptical distributions. Comput. Stat. to appear.

[10] Dunnett, C.W., and Sobel, M. (1954). A bivariate generalization of Student’s $t$-distribution, with tables for certain special cases. Biometrika 41, 153-169.

[11] Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange Multiplier tests in econometrics. In Z. Griliches and M.D. Intriligator (eds), “Handbook of Econometrics” volume 2, Elsevier, 775–826.

[12] Fama, E.F. (1965). The Behavior of Stock Market Prices. J. Bus. Univ. Chicago 38, 987-1007.

[13] Fama, E.F., and French, K.R. (1993). Common risk factors in the returns on stock and bonds. J. Financ. Econ. 33, 3-56.

[14] Fisher, R.A. (1925). Applications of “Student’s” distribution. Metron 5, 90-104.

[15] Garel, B., and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. Ann. Inst. Statist. Math. 47, 551-579.

[16] Hagerman, R.L. (1978). More evidence on the distribution of security returns. J. Finance 33, 1213–1221.

[17] Hájek, J. (1970). A characterization of limiting distributions of regular estimates. Z. Wahrsch. verw. Gebiete 14, 323-330.

[18] Hallin, M., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape: I. Optimal rank-based tests for sphericity. Ann. Statist. 34, 2707-2756.

[19] Helmert, F.R. (1875). Über die Bestimmung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. Z. Math. Phys. 20, 300-303.

[20] Henze, N. (1994). On Mardia’s kurtosis test for multivariate normality. Comm. Statist. Theor. Meth. 23, 1031–1045.

[21] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. Ann. Statist. 13, 1163-1174.

[22] Kan, R., and Zhou, G. (2003). Modeling non-normality using multivariate $t$: Implications for asset pricing. Technical report, Rotman School of Management, University of Toronto, Canada.
[23] Kotz, S., and Nadarajah, S. (2004). *Multivariate t Distributions and their Applications*. Cambridge University Press, USA.

[24] Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**, 112-133.

[25] Lange, K.L., Little, R.J.A., and Taylor, M.G. (1989). Robust Statistical Modeling Using the t Distribution. *J. Amer. Statist. Assoc.* **84**, 881-896.

[26] Le Cam, L. (1960). Locally asymptotically normal families of distribution. *Univ. Calif. Publ. in Stat.* **3**, 37-98.

[27] Le Cam, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35**, 1419-1455.

[28] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. Springer-Verlag, New York.

[29] Le Cam, L., and Yang, G.L. (2000). *Asymptotics in Statistics. Some Basic Concepts*. 2nd ed. Springer-Verlag, New York.

[30] Ley, C., and Paindaveine, D. (2010). On Fisher Information Matrices and Profile Log-Likelihood Functions in Generalized Skew-Elliptical Models. *Metron, special Issue on "Skew-symmetric and flexible distributions"* **68**, 235-250.

[31] Liu, C., and Rubin, D.B. (1995). ML estimation of the multivariate t distribution. *Statist. Sinica* **5**, 19-39.

[32] Lüroth, J. (1876). Vergleichung von zwei Werten des wahrscheinlichen Fehlers. *Astron. Nachr.* **87**, 209-220.

[33] Mandelbrot, B.B. (1963). The Variation of Certain Speculative Prices. *J. Bus. Univ. Chicago* **36**, 394-419.

[34] Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

[35] McElroy, T. (2007). Tail Index Estimation for Parametric Families Using Log Moments. Research report.

[36] Möttönen, R.J., and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparam. Statist.* **5**, 201-213.

[37] Pearson, K. (1895). Contributions to the mathematical theory of evolution, II. Skew variation in homogeneous material. *Philos. T. Roy. Soc. Lond. A* **186**, 343-414.

[38] Peel, D., and McLachlan, G.J. (2000). Robust mixture modelling using the t distribution. *Stat. Comput.* **10**, 339-348.

[39] Perry, P.R. (1983). More evidence on the nature of the distribution of security returns. *J. Financ. Quant. Anal.* **18**, 211-221.

[40] Pickands, J. III (1975). Statistical inference using extreme-order statistics. *Ann. Statist.* **3**, 119–131.

[41] Richardson, M., and Smith, T. (1993). A test for multivariate normality in stock returns. *J. Bus. Univ. Chicago* **66**, 295-321.
[42] Student (1908). The probable error of a mean. *Biometrika* **6**, 1-25.

[43] Swensen, A.R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16**, 54-70.

[44] Tyler, D.E. (1987). A distribution-free $M$-estimator of multivariate scatter. *Ann. Statist.* **15**, 234-251.