

UCLA

Department of Statistics Papers

Title

Pseudo Maximum Likelihood Estimation and Elliptical Distribution Theory in a Misspecified Model

Permalink

<https://escholarship.org/uc/item/96q113sn>

Authors

Yutaka Kano
Maia Berkane
Peter Bentler

Publication Date

2011-10-24

Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification

Maia Berkane

University of California, Los Angeles, CA, USA

Yutaka Kano

University of Osaka Prefecture, Osaka, Japan

Peter M. Bentler

University of California, Los Angeles, CA, USA

Received September 1992

Revised February 1993

Abstract: Recently, robust extensions of normal theory statistics have been proposed to permit modeling under a wider class of distributions (e.g., Taylor, 1992). Let X be a $p \times 1$ random vector, μ a $p \times 1$ location parameter, and V a $p \times p$ scatter matrix. Kano et al. (1993) studied inference in the elliptical class of distributions and gave a criterion for the choice of a particular family within this class to best describe the data at hand when the latter exhibit serious departure from normality. In this paper, we investigate the criterion for a simple but general set-up, namely, when the operating distribution is multivariate t with ν degrees of freedom and the model is also a multivariate t -distribution with α degrees of freedom. We compute the exact inefficiency of the estimators of μ and V based on that model and compare it to the one based on the multivariate normal model. Our results provide evidence for the choice of $\nu = 4$ proposed by Lange et al. (1989). In addition, we give numerical results showing that for fixed ν , the inflation of the variance of the pseudo maximum likelihood estimator of the scatter matrix, as a function of the hypothesized degrees of freedom α , is increasing in its domain.

Keywords: Quasi likelihood; Pseudo likelihood; Inflation; Elliptical distribution; Misspecification; Inefficiency

1. Introduction

Elliptical distribution theory has become increasingly popular in recent years (e.g., Fang and Anderson, 1990; Fang et al., 1990; Fang and Zhang, 1990). An

Correspondence to: M. Berkane, Department of Psychology, UCLA, Los Angeles, CA 90024-1563, USA.

important application has been to robustness research where it is used to evaluate the robustness of normal theory based statistics. Muirhead and Waterneaux (1980) investigated the robustness of a wide range of likelihood ratio test statistics for the covariance matrix of random vectors under the normality assumption when the true distribution is in fact elliptical. Their results show that under some regularity conditions, most of the computer outputs from existing software based on multinormal theory are still valid under a kurtosis adjustment. However, the estimators of the parameters of interest are not always efficient especially if the true distribution has a large kurtosis parameter, as pointed out by Tyler (1983).

Let X be a $p \times 1$ random vector having an elliptical distribution. Several authors have assumed a particular family of elliptical distributions as an alternative to the multinormal one. Let $u = (x - \mu)'V^{-1}(x - \mu)$, where μ is the $p \times 1$ location parameter vector and V is the $p \times p$ scatter matrix. Box and Tiao (1973) used the power exponential family (in the univariate case) defined by

$$f_x(x) = c_1 \exp\left(\frac{1}{2}|u|^{1/(1+\nu)}\right),$$

where ν is an extra parameter to be estimated and c_1 is a normalizing constant. Although this family can cover both leptokurtic and platykurtic distributions, its use is not computationally convenient when all parameters (μ, V, ν) are estimated simultaneously by the method of maximum likelihood as commented by Lange et al. (1989). These authors recommended a multivariate t -distribution

$$f_x(x) = c_2 \left(1 + \frac{u}{\nu}\right)^{-(p+\nu)/2}$$

and they suggest an apriori value of $\nu = 4$. In a recent paper, Taylor (1992) recommended estimating the extra kurtosis parameter in large samples, but fixing the parameter at "an appropriate" value in small samples. Kano et al. (1993) investigated properties of statistical inference using the likelihood method based on elliptical density functions

$$c|V|^{-1/2}f((x - \mu)'V^{-1}(x - \mu)|\nu), \quad (1.1)$$

where f is a known nonnegative function depending only on ν which may be multidimensional. They called ν the extra parameter vector. They proposed a criterion for the choice of a particular family of elliptical distributions for modelling data known to have come from an elliptical population with density

$$c_p|\tilde{V}_0|^{-1/2}g((x - \tilde{\mu}_0)'\tilde{V}_0^{-1}(x - \tilde{\mu}_0)), \quad (1.2)$$

where $\tilde{\mu}_0$ and $\tilde{V}_0$ are the true location and scatter, c_p a normalizing constant and g is a nonnegative function which may not be a member of (1.1). In this paper, we investigate in detail that criterion when the true distribution is the multivariate t with ν degrees of freedom, and the model is also a multivariate t -distribution but with degrees of freedom α , considered as fixed or cheaply estimated.

2. Quasi likelihood and pseudo likelihood

Let $\theta_1 = R^p$ and θ_2 the space of $p \times p$ positive definite matrices. For fixed family of elliptical densities and fixed nuisance (or extra) parameter ν in (1.1), one would like to find the density that is the closest to the true one in some sense. Kano et al. chose the Kullback–Leibler distance criterion defined as

$$E\left[\log |V|^{-1/2} f((x - \mu)' V^{-1} (x - \mu) | \nu)\right],$$

the expectation being taken with respect to the true density $g(x, \mu_0, V_0)$, and found the maximizer (μ_0, V_0) over all (μ, V) in $\theta_1 \times \theta_2$, of the above expression. (μ_0, V_0) is called the quasi true parameter (QTP), see, e.g., Nishii (1988). Hampel et al. (1986) found the QTP to be of the form $(\tilde{\mu}_0, s\tilde{V}_0)$ where s does not depend on $(\tilde{\mu}_0, \tilde{V}_0)$ but mainly on ν ; conditions for its uniqueness were also discussed. Since the quasi true parameter depends on the unknown function g , it is not known itself, and one would wish to estimate it using the maximum likelihood technique. Denote the first and second derivatives of f with respect to $u = (x - \mu)' V^{-1} (x - \mu)$ by f_u and f_{uu} respectively and evaluate these quantities at the quasi true parameter $\tilde{u}_0 = (x - \tilde{\mu}_0)' \tilde{V}_0^{-1} (x - \tilde{\mu}_0)$. In what follows, all expectations are taken with respect to g .

For a sample of size n , the likelihood is constructed in terms of μ and V based on the model (1.1) as

$$\prod_{k=1}^n |V|^{-1/2} f((x_k - \mu)' V^{-1} (x_k - \mu) | \hat{\nu}), \quad (2.1)$$

where $\hat{\nu}$ converges in probability to ν_0 or is fixed to some specified value. This likelihood is maximized with respect to (μ, V) . The obtained statistic $(\hat{\mu}, \hat{V})$ is known as the pseudo maximum likelihood estimator (PMLE), see, e.g., Gong and Sameniego (1981). Kano et al. (1993) studied the asymptotic properties of the PMLE and showed, under some regularity conditions, that

$$(a) \sqrt{n}(\hat{\mu} - \mu_0) \rightarrow N(0, b_{11}(b_{11} - a_{11})^{-2} V_0),$$

$$(b) \sqrt{n}(h(\hat{V}) - h(V_0)) \rightarrow N(0, \frac{1}{4} b_{22}(b_{22} - a_{22})^{-2} h_v \Omega h_v'),$$

where $h(V)$ is an $r \times 1$ vector valued, continuously differentiable function satisfying $h(cV) = h(V)$ for any constant $c > 0$, and h_v is the $r \times p^*$ ($p^* = \frac{1}{2}p(p+1)$) jacobian matrix, assumed to have full row rank, $\Omega = 2K_p'(V_0 \otimes V_0)K_p$, $\otimes$ is the right Kronecker product of matrices, and K_p is the $p^2 \times p^*$ transition matrix. Let $u_0 = (x - \mu_0)' V_0^{-1} (x - \mu_0)$, then

$$a_{11} = 2E\left(\frac{f_u}{f}\right) + \frac{4}{p}E\left(u_0 \frac{f_{uu}}{f}\right), \quad (2.2a)$$

$$a_{22} = \frac{1}{p(p+2)}E\left(u_0^2 \frac{f_{uu}}{f}\right) - \frac{1}{4} \quad (2.2b)$$

$$b_{11} = \frac{4}{p} E \left(u_0 \left(\frac{f_u}{f} \right)^2 \right), \quad (2.2c)$$

$$b_{22} = \frac{1}{p(p+2)} E \left(u_0^2 \left(\frac{f_u}{f} \right)^2 \right). \quad (2.2d)$$

The authors emphasized the choice of the elliptical model (1.1) based on the asymptotic variance of the PMLE. They suggested picking the model for which this asymptotic variance is the smallest. In (b) this device reduces the asymptotic variance to the minimum of $b_{22}(b_{22} - a_{22})^{-2}$ since the term $h_v \Omega h'_v$ is model free. We shall call $b_{22}(b_{22} - a_{22})^{-2}$ the inflation of the variance factor. In (a), we cannot ignore V_0 since it depends on the model. In what follows, we investigate the above quantities for a simple set-up.

3. Model

Assume the observed random vector $x(p \times 1)$ comes from an elliptical distribution with density $c_p g((x - \mu)' V^{-1} (x - \mu))$ with location μ and scatter matrix V . Let $u = (x - \mu)' V^{-1} (x - \mu)$. It is known (see for example Tyler, 1983) that the density of u is then

$$c_p \pi^{p/2} \Gamma^{-1}(\frac{1}{2}p) u^{p/2-1} g(u).$$

Assume the true density is the density of a multivariate t -distribution with ν degrees of freedom, i.e., that $g(u)$ has the form

$$g(u) = \left(1 + \frac{u}{\nu} \right)^{-(p+\nu)/2},$$

and take as a model the multivariate t -distribution with some degrees of freedom α , that is fixed a priori. Thus

$$f(u) = \left(1 + \frac{u}{\alpha} \right)^{-(p+\alpha)/2}. \quad (3.1)$$

Such kind of misspecification could very well be avoided by consistently estimating ν . In fact, for the t -distribution, the Mardia's multivariate kurtosis parameter $\beta_{2,p}$ is related to ν (Tyler, 1983) by

$$\beta_{2,p} = p(p+2) \left(\frac{\nu-2}{\nu-4} \right).$$

A consistent estimator of $\beta_{2,p}$ is given in Mardia (1970) as $1/n \sum_{i=1}^n ((x_i - \bar{x})' S^{-1} (x_i - \bar{x}))^2$ leading to a consistent estimator of ν . However, ν is sometimes assumed to have some fixed value α . In Lange et al., $\alpha = 4$. Since taking a fixed α is much cheaper and faster than estimating ν , it is relevant to investigate the

implications of such a strategy on the asymptotic variance of the PMLE. We propose to compute this asymptotic variance. We have that

$$f(u) = \left(1 + \frac{u}{\alpha}\right)^{-(p+\alpha)/2}, \quad \frac{f_u}{f} = -\frac{p+\alpha}{2\alpha} \left(1 + \frac{u}{\alpha}\right). \quad (3.2)$$

It is worth noticing that

$$\frac{f_{uu}}{f} = \left(1 + \frac{2}{p+\alpha}\right) \left(\frac{f_u}{f}\right)^2. \quad (3.3)$$

This relation may not be true in general for other elliptical distributions, but it will simplify future computations. Since $\mu_0 = \tilde{\mu}_0$ and $V_0 = s\tilde{V}_0$, and with u_0 replaced by $s^{-1}u$ where u depends on the true parameters, we are now ready to evaluate the quantities (2.2). We shall start with a_{22} and b_{22} since we first want to derive the asymptotic variance of $\hat{V}$.

$$\begin{aligned} E\left(u_0 \frac{f_u}{f}\right)^2 &= \left(\frac{p+\alpha}{2\alpha}\right)^2 \frac{\nu^{-p/2}}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} s^{-2} \int_0^\infty u^{p/2+1} \left(1 + \frac{u}{s\alpha}\right)^{-2} \left(t + \frac{u}{\nu}\right)^{-(p+\nu)/2} du, \end{aligned} \quad (3.4)$$

where $\beta(\frac{1}{2}\nu, \frac{1}{2}p)$ is the beta function. After we make the change of variable $t^{-1} = 1 + u/\nu$, the right-hand side of (3.4) becomes

$$\left(\frac{p+\alpha}{2\alpha}\right)^2 \frac{\nu^2}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} s^{-2} \int_0^1 t^{\nu/2-1} (1-t)^{p/2+1} \left(t + (1-t)\frac{\nu}{s\alpha}\right)^{-2} dt. \quad (3.5)$$

The scalar s which depends on the model, thus on α , plays an important role in the evaluation of the above integral. From Hampel et al. (1986, formula (5.3.13)), s must satisfy

$$E\left(\frac{u}{s} \frac{f_u\left(\frac{u}{s}\right)}{f\left(\frac{u}{s}\right)}\right) + \frac{1}{2}p = 0. \quad (3.6)$$

Marrona (1976) showed under certain mild assumptions on the operating distribution, the above equation enjoys a unique solution for each $\alpha > 0$. The equation (3.6) reduces to

$$E\left(\frac{u(p+\alpha)}{p\alpha} \left(1 + \frac{u}{s\alpha}\right)^{-1}\right) = s \quad (3.7a)$$

or

$$E\left(\frac{s}{s\alpha + u}\right) = \frac{1}{p+\alpha}. \quad (3.7b)$$

The following results will give some clarifications about the behaviour of s .

Proposition 1. Assume the distribution of u does not vanish and $E(u^{-1}) < \infty$, then the function $s(\alpha)$ implicitly defined by (3.6) is differentiable and $ds(\alpha)/d\alpha > 0$, for all $\alpha > 0$.

Note that Proposition 1 holds for a quite general family of operating distributions including the multivariate t -distribution as a special case.

Proposition 2. Let the dimension p of x be greater than 2 and let the operating distribution be the multivariate t -distribution with $\nu > 2$ degrees of freedom. Then we have:

- (i) $\alpha > \nu$, then $s > 1$ (so that $\nu/s\alpha < 1$),
- (ii) $\alpha = \nu$, then $s = 1$,
- (iii) $\alpha < \nu$, then $s < 1$ (so that $s\alpha(\nu) < 1$).

Further,

$$\lim_{\alpha \rightarrow \infty} s(\alpha) = \frac{\nu}{\nu - 2}.$$

Proof of Proposition 1. The integrability of $E(u^{-1})$ enables us to formally differentiate both sides of (3.7b) in terms of α , which results in

$$\frac{ds(\alpha)}{d\alpha} E\left(\frac{u}{(s\alpha + u)^2}\right) = E\left(\frac{s^2}{(s\alpha + u)^2}\right) - \frac{1}{(p + \alpha)^2}. \quad (3.8)$$

The coefficient of $ds(\alpha)/d\alpha$ is positive since u is not degenerate, hence the differentiability of $s(\alpha)$ holds. Applying the Schwarz inequality to the integral in (3.7b), we have that the right side of (3.8) is strictly positive, thus the proof is complete. $\square$

Proof of Proposition 2. We first note that $\nu > 2$ and $p > 2$ guarantees that $E(u) < \infty$ and that $E(u^{-1}) < \infty$. Thus Proposition 1 can apply. When $\alpha = \nu$, the model is correctly specified and the quasi true parameter must be equal to the true parameter. It follows that $s = 1$. Proposition 1 yield that $s(\alpha)$ is monotonically increasing. These facts prove the statements in (i)–(iii). From (3.7a) we establish

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} s(\alpha) &= \lim_{\alpha \rightarrow \infty} E\left(\frac{u(p + \alpha)}{p\alpha} \left(1 + \frac{u}{s\alpha}\right)^{-1}\right) \\ &= E\left(\lim_{\alpha \rightarrow \infty} \frac{u(p + \alpha)}{p\alpha} \left(1 + \frac{u}{s\alpha}\right)^{-1}\right) \\ &= E\left(\frac{u}{p}\right) = \frac{\nu}{\nu - 2}. \end{aligned}$$

The exchangeability of the limit and integral in the above is valid since $E(u) < \infty$, and the last equality holds for multivariate t -distributions. $\square$

In (3.5) consider three cases:

(i) $\alpha > \nu$. Write $\nu/s\alpha = 1 - \delta$ with $0 < \delta < 1$ then

$$\left(t + (1-t)\frac{\nu}{s\alpha}\right)^2 = (1 + \delta(1-t))^{-2} = \sum_{k=0}^{\infty} (k+1)\delta^k(1-t)^k,$$

and (3.5) becomes

$$\left(\frac{p+\alpha}{2\alpha}\right)^2 \frac{\nu^2}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} s^{-2} \sum_{k=0}^{\infty} (k+1)\delta^k \int_0^1 t^{\nu/2-1}(1-t)^{p/2+k+1} dt.$$

Since each integrand in the above summation is $\beta(\frac{1}{2}\nu, \frac{1}{2}p+k+2)$, it follows that

$$b_{22} = \frac{1}{p(p+2)} \left(\frac{p+\alpha}{2\alpha}\right)^2 \frac{\nu^2}{s^2\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \sum_{k=0}^{\infty} (k+1)\delta^k \beta(\frac{1}{2}\nu, \frac{1}{2}p+k+2). \quad (3.9)$$

Furthermore, because of (3.3), $a_{22} = (1 + 2/(p+\alpha))b_{22} - \frac{1}{4}$. Hence, for fixed ν , the inflation of the variance of the PMLE of the scatter matrix is $I_v(\alpha) = \frac{1}{4}b_{22}(\frac{1}{4} - 2b_{22}/(p+\alpha))^{-2}$.

(ii) $\alpha = \nu$. The model is correctly specified and (3.5) becomes

$$\left(\frac{p+\nu}{2}\right)^2 \frac{1}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \int_0^1 t^{\nu/2-1}(1-t)^{p/2+1} dt,$$

leading to $b_{22} = (p+\nu)/4(p+\nu+2)$ and $\frac{1}{4}b_{22}(b_{22} - a_{22})^{-2} = (p+\nu+2)/(p+\nu)$, with agrees with Lange et al. (1989).

(iii) $\alpha < \nu$. Write $s\alpha/\nu = 1 - \delta$ with $0 < \delta < 1$ then

$$\left(t + (1-t)\frac{\nu}{s\alpha}\right)^2 = (1-\delta)^2(t(1-\delta) + 1-t)^{-2} = \sum_{k=0}^{\infty} (k+1)\delta^k t^k$$

and (3.5) becomes

$$\left(\frac{p+\alpha}{2}\right)^2 \frac{1}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \sum_{k=0}^{\infty} (k+1) \int_0^1 t^{\nu/2+k-1}(1-t)^{p/2+1} dt \delta^k$$

and

$$b_{22} = \frac{1}{p(p+2)} \left(\frac{p+\alpha}{2}\right)^2 \frac{1}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \sum_{k=0}^{\infty} (k+1)\beta(\frac{1}{2}\nu+k, \frac{1}{2}p+2)\delta^k. \quad (3.10)$$

The inflation of the variance can be derived as previously, yielding the same formula with b_{22} as obtained via (3.9).

As α becomes infinite, $s = \nu/(\nu - 2)$, and (3.5) becomes

$$\begin{aligned} \frac{(\nu - 2)^2}{4\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \int_0^1 t^{\nu/2-3} (1-t)^{p/2+1} dt &= \left(\frac{\nu - 2}{2} \right)^2 \frac{\beta(\frac{1}{2}\nu - 2, \frac{1}{2}p + 2)}{\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \\ &= \left(\frac{\nu - 2}{4} \right) \frac{p(p + 2)}{(\nu - 4)}. \end{aligned}$$

The inflation of the variance of the PMLE of the scatter matrix is then $(\nu - 2)/(\nu - 4)$. This is equal to $\kappa + 1$, where κ is related to the Mardia multivariate kurtosis parameter $\beta_{2,p}$ by $\beta_{2,p} = p(p + 2)(\kappa + 1)$.

The derivations of this section lead to the following:

Proposition 3. *When the model is a multivariate t -distribution with α degrees of freedom while the true distribution is multivariate t -distribution with ν degrees of freedom, the cost of misspecifying ν , to the variance of the PMLE of the scatter matrix V for the three different situations is given in the following table.*

$\alpha < \nu$	$\alpha = \nu$	$\beta < \nu$	$\alpha \rightarrow +\infty$
$s < 1$	$s = 1$	$s > 1$	$s = \frac{\nu}{\nu - 2}$
b_{22} in (3.10)	$b_{22} = \frac{p + \nu}{4(p + \nu + 2)}$	b_{22} in (3.9)	$b_{22} = \frac{\nu - 2}{4(\nu - 4)}$
$I_V(\alpha) = \frac{1}{4}b_{22} \left(\frac{1}{4} - \frac{2b_{22}}{p + \alpha} \right)^{-2}$	$I_V(\alpha) = \frac{p + \nu + 2}{p + \nu}$	$I_V(\alpha) = \frac{1}{4}b_{22} \left(\frac{1}{4} - \frac{2b_{22}}{p + \alpha} \right)^{-2}$	$I_V(\alpha) = \frac{\nu - 2}{\nu - 4}$

4. Behaviour of $I_V(\alpha)$

The derivative of $I_V(\alpha)$ has a complicated form and depends on s which is an implicit function of α . Therefore it is difficult to study the behaviour of $I_V(\alpha)$ analytically. However, numerical results (Section 6) show that for fixed ν , $I_V(\alpha)$ decreases to $(p + \nu + 2)/(p + \nu)$ then increases from that value while being asymptotic to the line $I = \nu/(\nu - 2)$.

5. Inflation of the variance of the PMLE of the mean

The asymptotic covariance matrix of the PMLE of μ is $I_\mu(\alpha) = b_{11}(b_{11} - a_{11})^{-2}V_0$ with b_{11} and a_{11} in (2.2). Since V_0 is equal to sV_0 we need only focus on the scalar $b_{11}(b_{11} - a_{11})^{-2}s$ to compare different models. Following the same path of computations as in Section 3, it can be shown that

(i) $\Leftrightarrow < \nu$.

$$b_{11} = \frac{(p + \alpha)^2}{p\alpha\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \sum_{k=0}^{\infty} (k + 1)\beta(\frac{1}{2}\nu + 1, \frac{1}{2}p + k + 1)\delta^k, \quad (5.1)$$

$$a_{11} = \left(1 + \frac{2}{p + \alpha} \right) b_{11} - \frac{(p + \alpha)s\alpha}{\nu\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \sum_{k=0}^{\infty} \beta(\frac{1}{2}\nu + k + 1, \frac{1}{2}p)\delta^k, \quad (5.2)$$

(ii) $\nu = \alpha$.

$$I_{\mu}(\alpha) = \frac{\nu + p + 2}{\nu + p}, \quad (5.3)$$

(iii) $\alpha > \nu$.

$$b_{11} = \frac{(p + \alpha)^2}{p\alpha^2\nu s\beta(\frac{1}{2}\nu, \frac{1}{2}p)} \sum_{k=0}^{\infty} (k + 1)\beta(\frac{1}{2}\nu + 1, \frac{1}{2}p + k + 1)\delta^k, \quad (5.4)$$

$$a_{11} = \left(1 + \frac{2}{p + \alpha}\right)b_{11} - \frac{p + \alpha}{\alpha} \sum_{k=0}^{\infty} \beta(\frac{1}{2}\nu + 1, \frac{1}{2}p + k)\delta^k. \quad (5.5)$$

(iv) $\alpha \rightarrow \infty$.

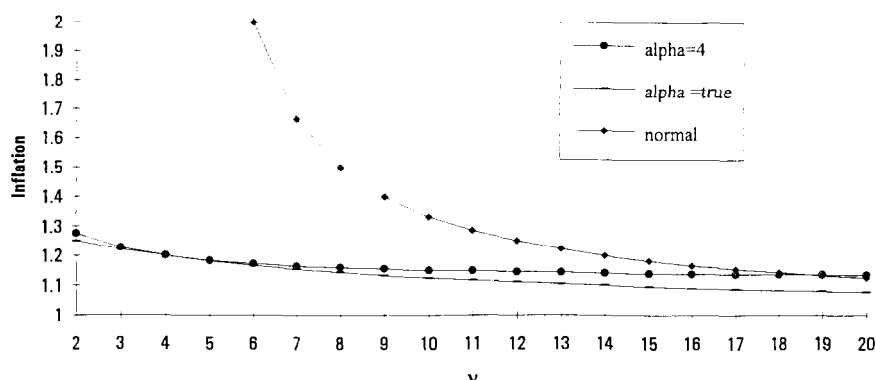
$$\lim_{\alpha \rightarrow \infty} I_{\mu}(\alpha) = 1 + \frac{2}{\nu - 2}. \quad (5.6)$$

6. Numerical results and Monte Carlo study

Table 1 gives the inflation of the PMLE of the variance for $\alpha = 4$ when ν ranges from 2 to 20 and p , the number of variables, is fixed at 6. For each value of ν , the scalar s is computed by use of the bisection method and then used in the expressions for a_{22} and b_{22} previously derived.

Table 1
Variance inflation of the scatter matrix when $\alpha = 4$

ν	$4^{-1}b_{22}(b_{22} - a_{22})^{-2}$	True	$(\nu - 2)/(\nu - 4)$
2	1.2758	1.2500	
3	1.2262	1.2222	
4	1.2000	1.2000	∞
5	1.1838	1.1818	3
6	1.1732	1.1666	2
7	1.1650	1.1538	1.6666
8	1.1591	1.1428	1.5000
9	1.1549	1.1333	1.4000
10	1.1507	1.1250	1.3333
11	1.1477	1.1176	1.2857
12	1.1451	1.1111	1.2500
13	1.1430	1.1052	1.2222
14	1.1411	1.1000	1.2000
15	1.1395	1.0952	1.1818
16	1.1381	1.0909	1.1666
17	1.1369	1.0869	1.1539
18	1.1358	1.0833	1.1428
19	1.1348	1.0800	1.1333
20	1.1339	1.0769	1.1250

Fig. 1. Graphs of $I_V(4)$, $I_V(\infty)$ and the true scalar.

According to the results of Table 1, for $2 < \nu < 20$, choosing the multivariate t -distribution with small degrees of freedom leads to less inflated variance of the PMLE of the scatter matrix than when choosing the normal model.

Fig. 1 gives the graphs of $I_V(4)$, $I_V(\infty)$ and the true scalar, for $2 < \nu < 20$. Table 2 gives $I_\mu(\alpha)$ and $I_V(\alpha)$ for a fixed value of ν and p , namely $\nu = 5$ and $p = 6$. They both increase and converge (slowly) to the normal model inflations.

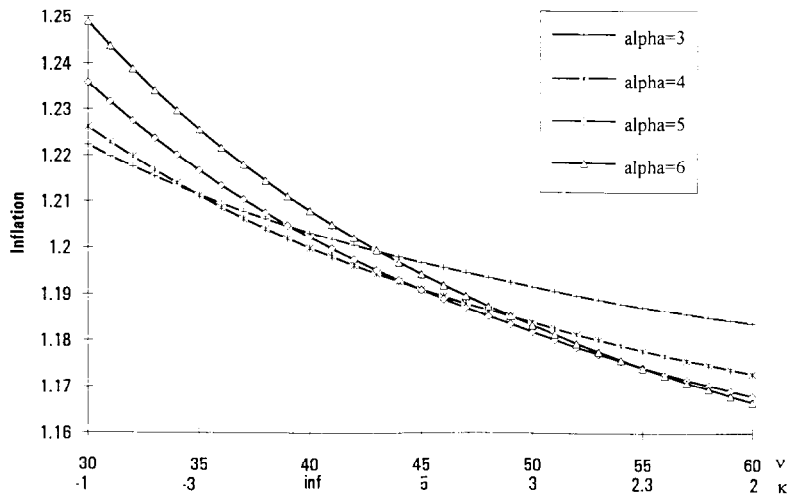
For $2 < \nu < 17$, the t -distribution with any arbitrary degrees of freedom leads to less inflated variance of the PMLE of the scatter matrix as compared to the standard normal model.

Table 2

Variance inflation of the mean vector and scatter matrix when $\nu = 5$

α	$b_{11}(b_{11} - a_{11})^{-2}$	$4^{-1}b_{22}(b_{22} - a_{22})^{-2}$
2	4.6072	1.2093
3	2.5296	1.1960
4	1.6621	1.1838
5	1.1818	1.1818
50	1.3885	1.4750
100	1.4723	1.6595
150	1.5143	1.7782
200	1.5402	1.8650
250	1.5579	1.9323
300	1.5709	1.9871
1000	1.6277	2.3234
2000	1.6437	2.4820
3000	1.6495	2.5596
4000	1.6525	2.6076
5000	1.6544	2.6392
6000	1.6557	2.6655
7000	1.6566	2.6843
8000	1.6573	2.6995

Note: Normal inflation of the variance of the mean is 1.6666, and for the variance is 3.

Fig. 2. Graphs of $I_V(3)$ to $I_V(8)$.

Since we have the following relation between ν and κ

$$\nu = 4 + \frac{2}{\kappa - 1},$$

$\nu = 4$ appears to be a good choice when κ is large (large departure from normality), as shown by Fig. 2, it leads to less inflated overall values, closer to the true ones, when these range from 2 to 6. This justifies the choice of $\nu = 4$ in Lange, et al. (1990).

We generate samples of size 15, 20, 50, 100 and 200 respectively, from the t -distribution with 5 degrees of freedom and from the normal distribution. For each sample, we estimate the degrees of freedom by use of the Mardia's coefficient of kurtosis and then compute the inflation of the variance of the

Table 3

Inflation of the variance of the PMLE of the variance due to estimating ν

ν	n	a.r.i. ^a
5	15	1.532
	20	1.322
	50	1.053
	100	1.014
	200	1.006
∞	20	1.002
	50	1.002
	100	1.002
	200	1.001

^a a.r.i. is the average ratio of the inflation of the variance of the PMLE of the variance to the true scalar.

PMLE of the variance. The average ratio of inflation to the true variance is reported in Table 3. For small sample size, estimating ν inflates the variance by an average factor of 1.53 for $n = 15$, decreasing to 1.006 for $n = 200$ when the true distribution is multivariate t with 5 degrees of freedom. Comparing this to 1.001 when the degrees of freedom is fixed at the value 4, increasing to 1.066 when it is fixed at 17, it seems more advantageous to fix ν at any value between 4 and 17 than to estimate it, when the sample size is small.

When the true distribution is multivariate normal with a specified covariance matrix, and the data is treated as multivariate t -distributed, the inflation of the variance due to estimating the extra parameter ν is very small, even for small samples. Fixing the degrees of freedom at some small value between 3 and 10 does not inflate the variance of the PMLE by much, in fact the inflation decreases from 1.142 to 1.050. This shows that even when the true distribution is normal, there is practically no cost in treating the distribution as multivariate t with specified (possibly small) degrees of freedom. This result agrees with the one given by Taylor (1992) for the inflation of the variance of estimators in the linear regression model with multivariate t -distributed errors.

Acknowledgement

This research was supported by the National Institute on Drug Abuse under grant DA01070. Technical assistance of Shinn Tzong Wu is greatly appreciated.

References

- Box, G.E.P. and G.C. Tiao, Bayesian inference in statistical analysis (Addison-Wesley, Reading, MA, 1973).
- Fang, K.-T. and T.W. Anderson, Statistical inference in elliptically contoured and related distributions (Allerton, New York, 1990).
- Fang, K.-T., S. Kotz and K.-W. Ng, Symmetric multivariate and related distributions (Chapman and Hall, London, 1990).
- Fang, K.-T. and Y.-T. Zhang, Generalized multivariate analysis (Springer-Verlag, Berlin, 1990).
- Gong, G., and F.J. Samaniego, Pseudo maximum likelihood estimation: Theory and application, *Ann. Statist.* **9** (1981) 861–869.
- Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, Robust statistics: The approach based on influence functions (John Wiley, New York, 1986).
- Kano, Y., M. Berkane and P.M. Bentler, Statistical inference based on pseudo maximum likelihood in elliptical population, *J. Amer. Statist. Assoc.*, **88** (1993) 135–143.
- Lange, K.L., R.J.A. Little and J.M.G. Taylor, Robust statistical modeling using the t -distribution, *J. Amer. Statist. Assoc.* **84** (1989) 881–896.
- Mardia, K.V., Measures of multivariate skewness and kurtosis with applications, *Biometrika* **57** (1970) 519–530.
- Marrona, R., Robust M-estimators of multivariate location and scatter, *Ann. of Statist.* **4** (1976) 51–67.
- Muirhead, R.J. and C.M. Waterneaux, Asymptotic distribution in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika* **67** (1980) 31–43.

- Nishii, R., Maximum likelihood principle and model selection when the true model is unspecified, *J. Multi. Anal.* **27** (1988) 392–403.
- Taylor, J.M.G., Properties of modelling the error distribution with an extra shape parameter, *Comp. Statist. & Data Anal.* **13** (1992) 33–46.
- Tyler, D.E., Robustness and efficiency properties of scatter matrices, *Biometrika* **70** (1983) 411–420.