Multivariate Analysis and Its Applications
IMS Lecture Notes - Monograph Series (1994) Volume 24

ON CLUSTER REGRESSION AND FACTOR ANALYSIS MODELS WITH ELLIPTIC t ERRORS

BY BRAJENDRA C. SUTRADHAR
Memorial University of Newfoundland

This paper deals with a two-stage cluster sampling problem. At the first stage k clusters are drawn at random, and then at the second stage n p-dimensional correlated observations are chosen under each cluster, which may be linearly related to certain covariates. The data of this type can be represented by a cluster regression model with suitable distributional assumption for the errors. In this paper, it is assumed that the error vector of the linear model follow an np-dimensional elliptically contoured t-distribution. Many elliptical distribution theory results are developed in the literature under the assumption that these n p-dimensional errors are uncorrelated. But, in the case of cluster sampling, errors are usually assumed to be equicorrelated. We, therefore, introduce a suitable $np \times np$ covariance matrix for n p-dimensional errors which takes the common intra-cluster correlation into account. We then study the likelihood inferences for the regression parameters (coefficients of the covariates) of the (linear) cluster regression model. Maximum likelihood estimators (m.l.e.) of the regression parameters are found to be more efficient than the generalized least squares estimators for smaller values of the degrees of freedom parameter of the elliptically contoured t-distribution. The asymptotic $(k \to \infty)$ distribution of the m.l.e. of the regression coefficients is also given.

Further, a factor analysis model is studied. Based on the assumption that n p-dimensional observations are uncorrelated and they follow np-dimensional elliptically contoured t-distribution, we have developed Neyman's partial score test for testing the suitability of the number of factors. The test statistic has asymptotically χ^2 distribution with suitable degrees of freedom. Moreover, the test is asymptotically optimal.

1. Introduction.

Elliptical distributions have been employed in two general approaches yielding somewhat different results. In one, p-dimensional

---

AMS 1980 Subject Classifications: 62H12, 62H15, 62H25, 62J05

Key words and phrases: Asymptotic chi-square distribution; clustered data; efficiency of the estimators; elliptically contoured t-errors; maximum likelihood estimation; Neyman's partial score test; testing covariance matrix.

random variables $Y_1, \cdots, Y_j, \cdots, Y_n$ are regarded as being distributed according to an np-dimensional elliptical distribution having the p.d.f. of the form

$$|\Lambda|^{-\frac{n}{2}} g\left[\sum_{j=1}^{n}(y_j-\mu)'\Lambda^{-1}(y_j-\mu)\right], \tag{1.1}$$

where $Y_1, \cdots, Y_n$ are pairwise uncorrelated but not necessarily independent. This elliptical class of distributions (1.1) have been studied by many authors. The null robustness of certain test statistics for testing hypotheses regarding μ and Λ (or Σ) has been studied, among others, by Kariya and Eaton (1977), Dawid (1977), Chmielewski (1980), Fraser and Ng (1980), Jensen and Good (1981), Kariya (1981a,b), Anderson, Fang and Hsu (1986). The optimality robustness of certain tests for μ and Λ (or Σ) concerning uniformly most powerful invariance (UMPI), and locally best invariance (LBI) properties has been treated by Kariya and Eaton (1977), Kariya (1981) and Kariya and Sinha (1985), among others. Similarly, the non-null robustness of certain tests has been studied by Kariya and Sinha (1985), among others, for the elliptical class of distributions (1.1).

It is well-known that many important test statistics which are developed based on normality are not non-null robust for the elliptical class of distributions (1.1). For example, the classical F-test for testing linear regression, and discriminant criteria to classify an observation into one of some of the elliptical populations, are not non-null robust. Then the question arises: how the power properties of the test are affected when the data really comes from a non-normal elliptical population. A general answer to this question is not possible, since the power properties usually depend on the specific alternative distribution. As there are many situations where the distributions involved may have heavier tails than those of the normal distribution, Sutradhar (1988, 1990) has considered the sub-class of elliptical t-distribution (a special case of (1.1)) as a parent population of the data and have shown the effect of non-normality on the power properties of certain tests. More specifically, it has been shown in Sutradhar (1988) that the power of the classical F-test for testing the linear regression depends on the degrees of freedom of the t-distribution. Sutradhar (1990) has shown that the probabilities of misclassification based on the well-known Fisher's linear discriminant criterion are generally smaller than the normal based misclassification probabilities. Thus, if a sample really comes from a t-population with $\nu(<\infty)$ degrees of freedom, say, the evaluation of classical classification error rates by normal-based probabilities would unnecessarily make an experimenter more suspicious.

The estimation of the location and scale parameter for the elliptical class of distributions (1.1), as well as, certain exact sampling distributional properties of these estimators, have also been discussed by many authors, among whom are Anderson, Fang and Hsu (1986), Sutradhar and Ali (1989), Kelker (1970), Thomas (1970), Strawderman (1974), Srivastava and Bilodeau (1989).

In the second approach, $Y_1, \cdots, Y_n$ are regarded as being independent and identically distributed according to the p-variate elliptical distribution. Thus, the likelihood function for μ and Λ is given by

$$|\Lambda|^{-\frac{n}{2}} \prod_{j=1}^{n} g\{(y_j - \mu)\Lambda^{-1}(y_j - \mu)\}. \tag{1.2}$$

Under the joint distribution (1.2), a number of situations were found, where certain corrected tests (correction for kurtosis of the elliptical distribution) for testing hypotheses regarding Σ (covariance matrix) in particular, retain their normal theory results. See for example, Browne (1982, 1984), Tyler (1982, 1983), Shapiro and Browne (1987), Browne and Shapiro (1987), Muirhead (1982), Muirhead and Waternaux (1980), and Satorra and Bentler (1988).

In Section 2, we propose a general likelihood approach which includes the above two approaches as special cases. The outline of the paper is also given in the same section.

2. General Likelihood. Let $Y = (Y_1, \cdots, Y_j, \cdots, Y_n)$ be a sample of size n, where $Y_j = (Y_{j1}, \cdots, Y_{jh}, \cdots, Y_{jp})'$ is a p-dimensional vector. Also let $Y^* = (Y_1', \cdots, Y_j', \cdots, Y_n')'$ denote the np-dimensional vector formed by stacking the n p-dimensional variables $Y_1, \cdots, Y_j, \cdots, Y_n$. Now consider k independent np-dimensional observations $y_1^*, \cdots, y_i^*, \cdots, y_k^*$, where y_i^*, for all $i = 1, \cdots, k$, is the ith realization of Y^*. Suppose the joint density function of k np-dimensional observations $y_1^*, \cdots, y_k^*$ is given by

$$F = K^k |\Lambda|^{-\frac{nk}{2}} \prod_{i=1}^{k} g\big[(y_i^* - 1_n \otimes \mu_i)'(I_n \otimes \Lambda^{-1})(y_i^* - 1_n \otimes \mu_i); \nu\big], \tag{2.1}$$

where ν is a shape parameter (usually a suitable function of the kurtosis parameter). Assuming that the covariance matrix $\Sigma = h(\nu)\Lambda$ exists, where $h(\nu)$ is a positive scalar function of ν, one may then reparametrize (2.1) by replacing Λ with $h^{-1}(\nu)\Sigma$, which yields the likelihood function of the form

$$F^* = K^* |\Sigma|^{-\frac{nk}{2}} \prod_{i=1}^{k} g\big[(y_i^* - 1_n \otimes \mu_i)'(I_n \otimes h(\nu)\Sigma^{-1})(y_i^* - 1_n \otimes \mu_i); \nu\big]. \tag{2.2}$$

The likelihood function (2.2) has been used by Shapiro and Browne (1987), among others, for the case with $n = 1$, in the context of covariance structures analysis. Further, when $k = 1$ in (2.2), one deals with the np-dimensional elliptical contoured distributions [cf. Anderson, Fang and Hsu (1986)].

Recently, Sutradhar (1993) has used the likelihood function (2.2) for the sub-class of elliptical t-distributions with $n = 1$ and large k, in the context of testing the hypotheses concerning Σ. In this paper, we deal with a multivariate cluster regression model with elliptical t-errors with their joint p.d.f.

analogous to the form (2.2). The regression model and the inferences about the regression parameters are discussed in Section 3. In Section 4, we provide an asymptotically optimal test for the goodness of fit of the factor analysis model where the common factors and the errors have elliptical t-distributions.

3. Cluster Regression Model with Elliptical t Errors. Consider two-stage cluster sampling in which k clusters are drawn at random at the first stage, and then at the second stage, n p-dimensional elements are chosen under each cluster following an np-dimensional elliptically contoured t-distribution. Let $y_i = (y'_{i1}, \cdots, y'_{ij}, \cdots, y'_{in})'$ be a vector of n p-dimensional observations from the ith cluster on a response variable y, where $y_{ij} = (y_{ij1}, \cdots, y_{ijh}, \cdots, y_{ijp})'$, y_{ijh} being the hth ($h = 1, \cdots, p$) variate of the jth ($j = 1, \cdots, n$) observation under the ith ($i = 1, \cdots, k$) cluster. Also let $x_{ijh} = (x_{ijh1}, \cdots, x_{ijhc})$ be the associated values of c covariates $x_1, \cdots, x_c$ which may or may not have influence on y. For example, in a zoological study one may be interested to know the effect of $c = 2$ covariates (age and weight, say) of the *Haltica oleracea* flea beetles on their $p = 2$ response variates, namely, the length of elytra and the length of the second antennal joint (microus). Suppose n beetles are trapped in each of k independent geographical regions. The data of this type may be modelled as

$$y_i = X_i\beta + \varepsilon_i, \qquad i = 1, \cdots, k, \tag{3.1}$$

where

$$X_i = \begin{bmatrix} X_{i1} : p \times pc \\ \vdots \\ X_{ij} : p \times pc \\ \vdots \\ X_{in} : p \times pc \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_h \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{ij} \\ \vdots \\ \varepsilon_{in} \end{bmatrix},$$

with

$$X_{ij} = \begin{bmatrix} x'_{ij1} & 0' & \cdots & 0' \\ 0' & x'_{ij2} & \cdots & 0' \\ \vdots & \vdots & & \vdots \\ 0' & 0' & \cdots & x'_{ijp} \end{bmatrix}, \qquad \beta_h = \begin{bmatrix} \beta_{h1} \\ \beta_{h2} \\ \vdots \\ \beta_{hc} \end{bmatrix}, \qquad \varepsilon_{ij} = \begin{bmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijh} \\ \vdots \\ \varepsilon_{ijp} \end{bmatrix},$$

$0'$ being the $1 \times c$ null vector. It is customary in two-stage cluster analysis to assume that the observations in a cluster are equicorrelated. For example, in the context of the illustration on flea beetles, given above, it may be reasonable to assume that n beetles under a specific geographical region will be equicorrelated. This leads us to assume that ε_i in (3.1) has np-dimensional zero mean vector and $np \times np$ covariance matrix $R(\rho) \otimes \Sigma$, where $\otimes$ denotes the Kronecker product, and $R(\rho) = (1-\rho)I_n + \rho U_n$ is the $n \times n$ equi-correlation matrix, ρ being the common intra-cluster correlation, I_n and U_n being the $n \times n$ identity and unit matrices respectively. For other choices of the correlation matrix, the methodology developed in the paper for testing the covariance matrix has fairly immediate generalizations. We further assume that ε_i in (3.1) has the np-dimensional elliptically contoured t-distribution with ν degrees of freedom. Then the likelihood function for the linear model may be written as

$$F^*(\beta, \rho, \Sigma, \nu) = C(\nu, k, n, p)|R(\rho) \otimes \Sigma|^{-\frac{k}{2}} \prod_{i=1}^{k} \{q_i(\nu)\}^{-\frac{\nu+np}{2}}, \tag{3.2}$$

where $q_i(\nu) = 1 + (\nu-2)^{-1}(y_i - X_i\beta)'(R(\rho) \otimes \Sigma)^{-1}(y_i - X_i\beta)$, $C(\nu, k, n, p)$ is the normalising constant given by

$$C(\nu, k, n, p) = [\{(\nu-2)^{\frac{\nu}{2}}\Gamma(\nu+np)/2\}/\{\pi^{\frac{np}{2}}\Gamma(\nu/2)\}]^k, \qquad \nu > 2.$$

The likelihood function (3.2) is a special case of the likelihood function (2.2) of the np-dimensional elliptical contoured distributions. This t-model (3.1–3.2) accomodates the usual multinormal model by letting $\nu \to \infty$, but, does not include the multivariate cauchy model for which neither the mean nor the variance exists.

3.1. Maximum Likelihood Estimate of β.

3.1.1. Case of Known ν, ρ and Σ. Let $f^* = \log F^*$ denote the log-likelihood function, where F^* is the likelihood function given by (3.2). Then solving the pc score functions $\partial f^*/\partial\beta = 0$, one obtains the maximum likelihood estimate of β as the solution of

$$\beta = \left(\sum_{i=1}^{k} q_i^{-1} P_i\right)^{-1} \sum_{i=1}^{k} q_i^{-1} Q_i y_i, \tag{3.3}$$

where $q_i = 1 + (\nu-2)^{-1}(y_i - X_i\beta)'(R(\rho) \otimes \Sigma)^{-1}(y_i - X_i\beta)$ as in (3.2), $Q_i = X_i'[R(\rho) \otimes \Sigma]^{-1}$, $P_i = Q_i X_i$. Let $\hat{\beta}$ denotes the solution of (3.3) for β. For the normal case when $\nu \to \infty$, or for the single cluster ($k = 1$) case with any $\nu > 2$, $\hat{\beta}$ reduces to $\hat{\beta}_{\mathrm{GLS}} = \left(\sum_{i=1}^{k} P_i\right)^{-1} \sum_{i=1}^{k} Q_i y_i$, where $\hat{\beta}_{\mathrm{GLS}}$ denotes the well-known GLS (generalized least square) estimate of β for the linear model (3.1). The GLS estimate is unbiased for β and has covariance matrix given by $\left(\sum_{i=1}^{k} P_i\right)^{-1}$. In the following theorem, we examine the relative efficiency of the maximum likelihood estimate $\hat{\beta}$ to the GLS estimate $\hat{\beta}_{\mathrm{GLS}}$.

THEOREM 3.1. *Let V_1 denotes the $pc \times pc$ covariance matrix of $\hat{\beta}$ and V_2 denotes the $pc \times pc$ covariance matrix of $\hat{\beta}_{\mathrm{GLS}}$. Then $V_1^{-1}V_2 = \{\nu(\nu+np)/(\nu-2)(\nu+np+2)\}I_{pc}$, where I_{pc} is the $pc \times pc$ identity matrix.*

PROOF. By direct computation, it may be shown that

$$-E\left(\frac{\partial^2 f^*}{\partial\beta\partial\beta'}\right) = \{\nu(\nu+np)/(\nu-2)(\nu+np+2)\}\sum_{i=1}^{k} P_i.$$

Then the theorem follows from the fact that $V_2 = \left(\sum_{i=1}^{k} P_i\right)^{-1}$.

It is clear from the theorem that for large np such that $np \simeq (np+2)$, the relative efficiency of $\hat{\beta}$ to $\hat{\beta}_{\text{GLS}}$ is $\nu/(\nu-2)$. Thus, for small ν, the maximum likelihood estimate of β will be highly more efficient than the generalized least square estimate of β. For large ν, $\hat{\beta}$ and $\hat{\beta}_{\text{GLS}}$ are identical, which is obvious.

Notice that the iteration procedure to solve $\hat{\beta}$ from (3.3) usually requires an intial estimate of β. Suppose $\hat{\beta}_{\text{GLS}}$ is used for β initially in q_i to compute β by (3.3). Let $\hat{\hat{\beta}}$ be this estimate which is given by

$$\hat{\hat{\beta}} = \left(\sum_{i=1}^{k} \hat{q}_i^{-1} P_i\right)^{-1} \sum_{i=1}^{k} \hat{q}_i^{-1} Q_i y_i, \tag{3.4}$$

where $\hat{q}_i = 1 + (\nu-2)^{-1}(y_i - X_i\hat{\beta}_{\text{GLS}})'(R(\rho)\otimes\Sigma)^{-1}(y_i - X_i\hat{\beta}_{\text{GLS}})$. If the iteration procedure is discontinued and one uses $\hat{\hat{\beta}}$ as the final estimate, efficiency is lost to a greater extent, which is shown below.

3.1.1.1. Relative Efficiency of $\hat{\hat{\beta}}$ to $\hat{\beta}_{\text{GLS}}$. We express $\hat{\hat{\beta}}$ in (3.4) as

$$\hat{\hat{\beta}} = \beta + \left(\sum_{i=1}^{k} \hat{q}_i^{-1} P_i\right)^{-1} \left(\sum_{i=1}^{k} \hat{q}_i^{-1} Q_i \varepsilon_i\right), \tag{3.5}$$

where ε_i has np-dimensional elliptically contoured t-distribution with zero mean, $R(\rho)\otimes\Sigma$ covariance matrix and ν degrees of freedom. If $\hat{q}_i$ is assumed to be known, for example, $\hat{q}_i$ is replaced by $E(\hat{q}_i) = m_i$ (say), then $\hat{\hat{\beta}}$ would be less efficient than the Gauss-Markoff estimator $\hat{\beta}_{\text{GLS}}$. As it is shown below, also for unknown $\hat{q}_i$, the first step estimator $\hat{\hat{\beta}}$ is less efficient than $\hat{\beta}_{\text{GLS}}$.

Rewrite (3.5) as $\hat{\hat{\beta}} = \beta + g\ (\hat{q}_1, \cdots, \hat{q}_k, \varepsilon_1', \cdots, \varepsilon_k')$. Then an approximate formula, up to second order, for the variance of $\hat{\hat{\beta}}$ can be obtained by using the Taylor series expansion of $g(\hat{q}_1, \cdots, \hat{q}_k, \varepsilon_1', \cdots, \varepsilon_k')$ evaluated at $E(\hat{q}_i) = m_i$ and $E(\varepsilon_i') = 0' : 1 \times np$ for all $i = 1, \cdots, k$. After some algebra we obtain

$$\text{var}(\hat{\hat{\beta}}) \simeq \left(\sum_{i=1}^{k} m_i^{-1} P_i\right)^{-1} \left(\sum_{i=1}^{k} m_i^{-2} P_i\right) \left(\sum_{i=1}^{k} m_i^{-1} P_i\right)^{-1}, \tag{3.6}$$

where

$$\begin{aligned} m_i =& E(\hat{q}_i) \\ =& 1 + (\nu-2)^{-1}\left[\text{trace}\left\{I_{np} - 2W_{ii} + W_{ii}W_{ii}'\right\} + \sum_{\ell\neq i} \text{trace}\left\{W_{\ell i}W_{\ell i}'\right\}\right], \end{aligned}$$

with $W_{\ell i} = (R(\rho)\otimes\Sigma)^{-\frac{1}{2}}X_\ell\left(\sum_{i=1}^k P_i\right)^{-1}X_i'(R(\rho)\otimes\Sigma)^{-\frac{1}{2}}$, for all $\ell, i = 1,\cdots,k$.

Notice that when $m_i = m$ for all i, $\hat{\hat{\beta}}$ reduces to $\hat{\beta}_{\text{GLS}}$. Now to show that $\hat{\hat{\beta}}$ is less efficient than $\hat{\beta}_{\text{GLS}}$, it is sufficient to consider the case $m_1 = m \pm \delta$, $m_2 = \ldots = m_k = m$ and examine the change of $\text{Var}(\hat{\hat{\beta}})$ due to change in m_1 because of δ. It can be shown that

$$\text{Lt}_{\delta\to 0}\frac{\partial\ \text{Var}(\hat{\hat{\beta}})}{\partial\delta} = \left[\begin{array}{ll} 2(\Sigma P_i)^{-1}P_1\Big(\sum_{i=1}^k P_i\Big)^{-1}\Big(1-\frac{1}{m}\Big) & \text{when}\quad m_1 = m+\delta, \\ -2(\Sigma P_i)^{-1}P_1\Big(\sum_{i=1}^k P_i\Big)^{-1}\Big(1-\frac{1}{m}\Big) & \text{when}\quad m_1 = m-\delta. \end{array}\right.$$

Since $m \geq 1$ always and P_i's are positive definite, it follows that $\text{Var}(\hat{\hat{\beta}}_u) \geq V(\hat{\beta}_{u,\text{GLS}})$ for all $u = 1,\cdots,pc$, where $\hat{\hat{\beta}}_u$ and $\hat{\beta}_{u,\text{GLS}}$ are the uth components of $\hat{\hat{\beta}}$ and $\hat{\beta}_{\text{GLS}}$ respectively. As it is shown in Theorem 3.1, the final m.l.e. of β is, however, more efficient than $\hat{\beta}_{\text{GLS}}$.

3.1.2. Case of Unknown ν, ρ and Σ. In practice ν, ρ and Σ are rarely known and β must be estimated simultaneously with the estimation of ν, ρ and Σ. In this case, the maximum likelihood estimates $\hat{\beta}$, $\hat{\Sigma}$, $\hat{\rho}$ and $\hat{\nu}$ of β, Σ, ρ and ν respectively, are obtained by solving simultaneously the likelihood equations for Σ, ρ and ν along with the likelihood equation for β given in (3.3).

3.2. Asymptotic Distribution of $\hat{\beta}$. Using $\hat{\beta}_{\text{GLS}}$ as an initial estimate for β, the solution for β after first iteration is given by $\hat{\hat{\beta}}$ (3.4). By (3.5) it is easy to show that $\hat{\hat{\beta}}$ is also (similar to $\hat{\beta}_{\text{GLS}}$) unbiased for β. Using similar argument at every stage of the iteration procedure, it follows that $\hat{\beta}$ (m.l.e.) is unbiased for β. Further in (3.1), $y_1,\cdots,y_i,\cdots,y_k$ are independent vector observations. Now by applying Theorem 3.1 and the well-known central limit theorem, we can state the asymptotic behaviour of $\hat{\beta}$ as in the following theorem.

THEOREM 3.2. *The asymptotic distribution of the maximum likelihood estimator $\hat{\beta}$ is multivariate normal with mean vector β and covariance matrix $\{(\nu-2)(\nu+np+2)/\nu(\nu+np)\}\left(\sum_{i=1}^k X_i'[R(\rho)\otimes\Sigma]^{-1}X_i\right)^{-1}$, which can be consistently estimated by $\{(\hat{\nu}-2)(\hat{\nu}+np+2)/\hat{\nu}(\hat{\nu}+np)\}\left\{\sum_{i=1}^k X_i'[R(\hat{\rho})\otimes\hat{\Sigma}]^{-1}X_i\right\}^{-1}$.*

4. Factor Analysis Models. Consider the factor analysis model

$$y_{ij} = X_{ij}\beta + \Lambda z + e_{ij}, \tag{4.1}$$

where y_{ij} is the p-dimensional jth observation ($j = 1, \cdots, n$) in the ith ($i = 1, \cdots, k$) cluster, X_{ij} is the $p \times pc$ design matrix, β is the $pc \times 1$ regression coefficient vectors as in Section 3, Λ is a $p \times m$ matrix of unknown factor loadings, z is the $m \times 1$ vector of unobservable common factors and e_{ij} is a $p \times 1$ vector of errors or residuals. In classical factor analysis, a maximum likelihood estimation factor analysis is carried out under the assumption that each of z and e_{ij} has multinormal distribution. Instead of normality, it is now assumed that $y_{i1}, \cdots, y_{ij}, \cdots, y_{in}$ (i.e. $n \geq 1$ observations in the ith cluster) have the joint np-dimensional elliptical t-distribution

$$f(y_{i1}', \cdots, y_{in}') = K(\nu, n, p)|I_n \otimes \Sigma|^{-\frac{1}{2}}\{q_i(\nu)\}^{-\frac{\nu+np}{2}}, \tag{4.2}$$

where $q_i(\nu) = 1+(\nu-2)^{-1}$ trace $\{\Sigma^{-1}\sum_{j=1}^{n}(y_{ij}-X_{ij}'\beta)(y_{ij}-X_{ij}'\beta)'\}$ with $\Sigma = \Lambda\Lambda' + \psi$, where $E(e_{ij}e_{ij}) = \psi = \text{Diag}(\psi_1, \cdots, \psi_p)$. Note that unlike Section 3, n observations of the ith cluster have been assumed to be uncorrelated in model (4.2). For the case with $n = 1$, that is, when each cluster contains a single p-dimensional observation, and when $f(\cdot)$ is the p.d.f. of a general elliptical distribution of the form (2.2), Browne and Shapiro (1987) have proposed a correction to the normal based minimum discrepancy test for testing the null hypothesis $\Sigma = \Sigma_0$, where $\Sigma_0 = \widehat{\Lambda}\widehat{\Lambda}' + \hat{\psi}$, Σ being the covariance matrix of the elliptical distribution, $\widehat{\Lambda}$ and $\hat{\psi}$ being the suitable estimates of Λ and ψ respectively. See also Satorra and Bentler (1988) for a similar but different scale correction to the normal theory maximum likelihood discrepancy test. Their correction takes the kurtosis of the elliptic distribution into account, and, in practice, the kurtosis parameter involved in the corrected statistic is replaced by its consistent estimate.

We remark here that although Browne and Shapiro's result is developed for general elliptic distribution, in practice we usually encounter with normal elliptic or t-elliptic or cauchy elliptic data. Our as well as Browne and Shapiro's results are not useful for cauchy elliptic data since we assume the existence of the covariance matrix. Otherwise, the corrected test statistic of Browne and Shapiro (1987) has asymptotically χ^2 null distribution. But, in general, the power properties of this test is not known.

In the following section, we develop the score test due to Neyman (1959). There are two main reasons to propose this test. First, this test is asymptotically locally optimal, and also asymptotically equivalent to the likelihood ratio and Wald's tests (cf. Moran (1970)). Second, the likelihood ratio and Wald's tests require the maximum likelihood estimates of the parameters which may be cumbersome to obtain by solving the likelihood equations mentioned in Section 3.1.2. Unlike the likelihood ratio and Wald's tests, the score test, however, requires only $\sqrt{k}$-consistent estimates for the nuisance parameters, which need not necessarily be the maximum likelihood estimates.

4.1. Neyman's Partial Score Test for $\Sigma = \Sigma_0$. Let $\sigma = \text{vec}\,(\Sigma)$, where vec (Σ) denotes the $\{p(p+1)/2\} \times 1$ vector formed by stacking the distinct elements of Σ. Also let D_σ be the $p \times p$ symmetric matrix obtained from the likelihood equation (3.2) for Σ after putting $\rho = 0$. Then, for $\ell \geq h$, $h, \ell = 1, \cdots, p$, the $p(p+1)/2$ score functions for $\sigma_{11}, \cdots, \sigma_{h\ell}, \cdots, \sigma_{pp}$ are computed by stacking the distinct elements of the $p \times p$ symmetric matrix D_σ. Let d_σ denote this $\{p(p+1)/2\} \times 1$ score vector. Similarly, we construct pc score functions for the elements of β after putting $\rho = 0$ in the likelihood equation. Let this $pc \times 1$ score vector be denoted by d_β. Next the score function for ν, d_ν (say), is obtained from the likelihood function for ν after puting $\rho = 0$. Further, let β^* and ν^* be the consistent estimates for β and ν, which are not necessarily the maximum likelihood estimates. The derivation of these consistent estimators is provided in Section 4.1.2. Now for testing $\sigma = \sigma_0$ (σ_0 being the $\{p(p+1)/2\} \times 1$ vector formed by stacking the distinct elements of Σ_0) Neyman's partial score test statistic (cf. Neyman (1959)) λ^* (say) is given by

$$\begin{aligned}\lambda^* =& T' \left[D_{\sigma\sigma} - (D_{\sigma\beta}\; D_{\sigma\nu}) \begin{pmatrix} D_{\beta\beta} & D_{\beta\nu} \\ & D_{\nu\nu} \end{pmatrix}^{-1} \begin{pmatrix} D'_{\sigma\beta} \\ D'_{\sigma\nu} \end{pmatrix} \right]^{-1} \\ &\cdot\; T|_{\sigma=\sigma_0, \beta=\beta^*, \nu=\nu^*},\end{aligned} \tag{4.3}$$

where

$$T = d_\sigma - (D_{\sigma\beta}\; D_{\sigma\nu}) \begin{pmatrix} D_{\beta\beta} & D_{\beta\nu} \\ & D_{\nu\nu} \end{pmatrix}^{-1} \begin{pmatrix} d_\beta \\ d_\nu \end{pmatrix},$$

is the $pc \times 1$ residual vector evaluated at $\sigma = \sigma_0$, $\beta = \beta^*$, and $\nu = \nu^*$, and where, for example, $D_{\sigma\sigma}$ is the (σ, σ) section of the information matrix

$$D = \begin{bmatrix} D_{\sigma\sigma} & D_{\sigma\beta} & D_{\sigma\nu} \\ & D_{\beta\beta} & D_{\beta\nu} \\ & & D_{\nu\nu} \end{bmatrix}, \tag{4.4}$$

evaluated at $\sigma = \sigma_0$, $\beta = \beta^*$, and $\nu = \nu^*$, with

$$D_{\sigma\sigma} = -E\left(\frac{\partial^2 \log F^*}{\partial\sigma\partial\sigma'}\right) : \{p(p+1)/2 \times p(p+1)/2\},$$

F^* being the likelihood function obtained from (4.2). The other sections of the D matrix are defined similarly.

Under the null hypothesis $\sigma = \sigma_0$, the test statistics λ^* has approximately χ^2 distribution with $p(p+1)/2$ degrees of freedom. This test, at least approximately, is asymptotically unbiased in estimating a pre-assigned level of significance (cf. Bartoo and Puri (1967)). Also the test is locally asymptotically most powerful and asymptotically equivalent to the likelihood ratio and Wald's tests (cf. Moran (1970)).

4.1.1. Computation of D. Differentiation of d_σ with respect to σ vector and the direct evaluation of the expectation of the second derivatives based on the elliptical t-distribution (4.2) yields the $D_{\sigma\sigma}$ matrix in (4.4) as

$$D_{\sigma\sigma} = [d(1,1)\;\; d(1,2)\cdots d(h,\ell)\cdots d(p,p)], \tag{4.5}$$

where, for $\ell \geq h$, $h,\ell = 1,\cdots,p$, $d(h,\ell)$ is a $\{p(p+1)/2\}$-dimensional column vector formed by stacking the distinct elements of the $p \times p$ symmetric matrix

$$D_{h,\ell} = \frac{kn}{2}\left[\{\sigma^h \otimes (\sigma^\ell)'\} - \nu(\nu-2)^{-1}(\nu+np+2)^{-1} \text{ trace } \{\sigma^h \otimes (\sigma^\ell)'\}I_p\right],$$

where σ^h denotes the hth column of Σ^{-1}.

By similar calculations, we obtain

$$D_{\beta\beta} = -E\left[\frac{\partial}{\partial\beta}d_\beta\right] = \{\nu(\nu+np)/(\nu-2)(\nu+np+2)\}\sum_{i=1}^{k} P_i, \tag{4.6}$$

where $P_i = X_i{}'[I_n \otimes \Sigma]^{-1}X_i$, and

$$\begin{aligned} D_{\nu\nu} =& k\{\psi'(z_2)/4 - \psi'(z_1)/4 - (\nu-4)/2(\nu-2)^2\} \\ & - k\nu/2(\nu-2)\left[2/(\nu+np) - (\nu+2)/\{(\nu-2)(\nu+np+2)\}\right], \end{aligned}$$

where $\psi'(z_2)$ and $\psi'(z_1)$ are the derivatives of the digamma functions $\psi(z_2)$ and $\psi(z_1)$ with respect to z_2 and z_1 respectively. The (σ,β) section of the information matrix can be shown to be the null matrix of order $\{p(p+1)/2\} \times pc$. Also, $D_{\nu\beta}$ can be shown to be the null vector of dimension $p(p+1)/2$. Finally, the (σ,ν) section of the information matrix is formed by stacking the distinct elements of the $p \times p$ matrix

$$-\{1/2(\nu+np)\}\{1 - \nu/(\nu-2)(\nu+np+2)\}\sum_{i=1}^{k} P_i,$$

where P_i is the same matrix as in (4.6).

4.1.2. Consistent Estimates for Nuisance Parameters. In testing $\Sigma = \Sigma_0$, β and ν are considered to be the nuisance parameters. The ordinary least square (OLS) estimator, $\hat{\beta}_{\text{OLS}}$, may be shown to be a consistent estimator of β. This is simple to compute too. In the present set up,

$$\hat{\beta}_{\text{OLS}} = \left(\sum_{i=1}^{k} X_i{}'X_i\right)^{-1} \sum_{i=1}^{k} X_i{}'y_i,$$

with its variance given by

$$\text{var}\ (\hat{\beta}_{\text{OLS}}) = \left(\sum_{i=1}^{k} X_i{}'X_i\right)^{-1} \sum_{i=1}^{k} \{X_i{}'(I_n \otimes \Sigma)X_i\} \left(\sum_{i=1}^{k} X_i{}'X_i\right)^{-1},$$

where X_i is the $np \times pc$ design matrix and y_i is the np-dimensional observation vector. Writing $\sum_{i=1}^{k} X_i{}'X_i$ as $X'X$, where X is the $knp \times p$ matrix, one obtains

$$\begin{aligned} \text{trace}\ \{\text{var}\ (\hat{\beta}_{\text{OLS}}\} &= \text{trace}\ \left[(X'X)^{-1}X'(I_{kn} \otimes \Sigma)X(X'X)^{-1}\right] \\ &= \text{trace}\ \left[(I_{kn} \otimes \Sigma)X(X'X)^{-2}X'\right]. \end{aligned} \tag{4.7}$$

Under certain mild conditions, that is, when the largest eigenvalue of Σ is bounded, also when the smallest eigenvalue of $X'X$ is very large, one may show that [cf. Amemiya (1985, Chapter 6)] trace{ var $(\hat{\beta}_{\text{OLS}})$} converges to zero, showing that $\hat{\beta}_{\text{OLS}}$ is consistent for β. In the notation of Section 4.1, we therefore have $\beta^* = \hat{\beta}_{\text{OLS}}$.

In order to construct a consistent estimator ν^* for ν, one may proceed as follows. Following Mardia (1970), the multivariate measure of kurtosis of the elliptical t-distribution may be written as

$$\beta_2 = \int [(y_i - X_i\beta)'(I_n \otimes \Sigma^{-1})(y_i - X_i\beta)]\partial F^*, \tag{4.8}$$

where F^* is as in (3.2) after putting $\rho = 0$. By direct integration we obtain

$$\beta_2 = \left(\frac{\nu - 2}{\nu - 4}\right) f(n, \sigma),$$

where

$$\begin{aligned} f(n,\sigma) =& n\Big[3 \sum_{h=1}^{p} (\sigma^{hh})^2 (\sigma_{hh})^2 \\ &+ \sum_{h \neq h'}^{p} (\sigma_{h'h'})^2 \{\sigma^{hh}\sigma^{h'h'} + (\sigma^{hh'})^2\}\Big], \end{aligned} \tag{4.9}$$

and where $\sigma^{hh'}$ and σ_{hh} are the (h,h')th element of Σ^{-1} and Σ respectively. Then, by using

$$\beta_2^* = \frac{1}{k} \sum_{i=1}^{k} \left[(y_i - X_i\beta^*)'(I_n \otimes S^{-1})(y_i - X_i\beta^*)\right]^2, \tag{4.10}$$

a consistent estimator for β_2, one obtains the consistent estimator for ν as

$$\nu^* = 2[2\beta_2^* - f(n,s)]/[\beta_2^* - f(n,s)], \tag{4.11}$$

where $f(n,s)$ is computed by (4.9) after replacing Σ by S, S being the sample covariance matrix given by

$$S = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - X_{ij}\beta^*)(y_{ij} - X_{ij}\beta^*)'/(kn - pc).$$

5. Concluding Remarks. Elliptical distributions have been employed in two general approaches. In one, an $n \times p$ data matrix is regarded as being distributed according to an np-dimensional elliptical distribution. Elements in different rows (i.e. n p-dimensional observations) are regarded as uncorrelated but not independent. In the second approach rows of the data matrix are regarded as being independent and identically distributed according to a p-variate elliptical distribution. In the present paper we have provided a more general approach which contains the above two general approaches as special cases. More specifically, we have considered k independent clusters (groups), where each cluster has n p-dimensional observations regarded as being generated according to an np-dimensional elliptical distribution. It has been assumed that the $p \times p$ covariance matrix exists. Since in practice, we mostly encounter normal or t elliptic data (with suitable covariance matrix) and because distribution theory is well-developed for normal data, we have concentrated, in this paper, to elliptical t-data only. In Section 3, likelihood inference is given for the regression coefficients of a cluster regression model assuming that the rows of the $n \times p$ data matrix, that is, n observations in a cluster are equicorrelated. Section 4 deals with a factor analysis model based on the assumption that the rows of the $n \times p$ matrix in a cluster are uncorrelated. In contrast to Browne and Shapiro's (1987) discrepancy test (for testing the covariance matrix) we have discussed Neyman's partial score test which is asymptotically locally most powerful. A simulation study could be done to examine the large as well as small sample performance of the Neyman's partial score test compared to the Brown-Shapiro (1987), Satorra-Bentler (1988) and the normal theory based tests but such a study was not chosen in the present paper.

Acknowledgements. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada. The author thanks the referees for their comments on the previous version.

REFERENCES

ABRAMOWITZ M., and STEGUN, I. A. (Eds). (1965). *Handbook of Mathematical Functions.* New York: Dover.

AMEMIYA, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge, Massachusetts.

ANDERSON, T. W., FANG, K. T., and HSU, H. (1986). Maximum likelihood estimates and likelihood ratio criteria for multivariate elliptically contoured distributions. *Canadian Journal of Statistics* **14**, 55–59.

BARTOO, J. B., and PURI, P. S. (1967). On optimal asymptotic tests of composite statistical hypothesis. *Ann. Math. Statist.* **38**, 1845–1852.

BROWNE, M. W. (1982). Covariance structures, *Topics in Applied Multivariate Analysis*, ed. D. M. Hawkins, Cambridge, U.K.: Cambridge University Press, 72–41.

BROWNE, M. W. (1984). Asymptotically distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* **37**, 62–83.

BROWNE, M. W. and SHAPIRO, A. (1987). Adjustments for kurtosis in factor analysis with elliptically distributed errors. *J. R. Statist. Soc.* **B49**, 346–352.

CHMIELEWSKI, M. A. (1980). Invariance scale matrix hypothesis tests under elliptical symmetry. *Journal of Multivariate Analysis* **10**, 343–350.

DAWID, A. P. (1977). Spherical matrix distributions and a multivariate model. *J. R. Statist. Soc.* **B39**, 254–261.

FRASER, D. A. S., and NG, K. W. (1980). Multivariate regression analysis with spherical error, *Multivariate Analysis* 5 (P. R. Krishnaiah, ed). North-Holland, New York, 369–386.

JENSEN, D. R., and GOOD, I. J. (1981). Invariant distributions associated with matrix laws under structural symmetry. *J. R. Statist. Soc.* **B43**, 327–332.

JORESKOG, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32**, 443–482.

KARIYA, T. (1981). Robustness of multivariate tests. *Ann. Statist.* **9**, 1267–1275.

KARIYA, T. and EATON, M. L. (1977). Robust tests for spherical symmetry. *Ann. Statist.* **5**, 206–215.

KARIYA T. and SINHA, B. K. (1985). Non-null and optimality robustness of some tests. *Ann. Statist.* **13**, 1183–1196.

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya A* **32**, 419–430.

MARDIA, K. V. (1970). Measures of multivariate skewnness and kurtosis with applications. *Biometrika* **57**, 519–530.

MORAN, P. A. P. (1970). On asymptotically optimal tests of composite hypothesis. *Biometrika* **57**, 47–55.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

MUIRHEAD, R. J., and WATERNAUX, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for non-normal populations. *Biometrika* **67**, 31–43.

NEYMAN, J. (1959). Optimal asymptotic tests for composite hypotheses, In *Probability and Statistics*, Ed. U. Grenander, 213–234. Wiley, New York.

SATORRA, A., and BENTLER, P. M. (1988). Scalilng corrections for chi-square statistics in covariance structure analysis. In *Proc. Amer. Statist. Assoc.* 308–313.

SHAPIRO, A. and BROWNE, M. W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of American Statistical Association* **82**, 1092–1097.

SRIVASTAVA, M. S. and BILODEAU, M. (1989). Stein estimation under elliptical distributions. *Journal of Multivariate Analysis* **28**, 247–259.

STRAWDERMAN, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *Journal of Multivariate Analysis* **4**, 255–264.

SUTRADHAR, B. C. (1988). Testing linear hypothesis with t error variable. *Sankhya Series* **B50**, 175–180.

SUTRADHAR, B. C. (1990). Discrimination of observations into one of two t populations. *Biometrics* **46**, 827–835.

SUTRADHAR, B. C. (1993). Score test for the covariance matrix of the elliptical t distribution. *Journal of Multivariate Analysis* **46**, 1–2.

SUTRADHAR, B. C., and ALI, M. M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model. *Journal of Multivariate Analysis* **29**, 155–162.

THOMAS, D. H. (1970). Some contributions to radical probability distributions, statistics, and the operational calculi. Ph.D. Dissertation, Wayne

State University.

TYLER, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika* **69**, 429–436.

TYLER, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika* **70**, 411–420.

DEPARTMENT OF MATHEMATICS STATISTICS
MEMORIAL UNIVERSITY OF NEWFOUNDLAND
ST. JOHN'S NEWFOUNDLAND
CANADA A1C 5S7