International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 3 (Special Issue), 2004, pp 91–103

On the consistency of the estimators of the parameters of elliptically contoured distributions

Brajendra C. Sutradhar
Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, Newfoundland
Canada A1C 5S7
E-mail: bsutradh@math.mun.ca

[Received April 23, 2004; Accepted August 23, 2004]

Abstract

Elliptically contoured distributions have densities with equiprobable surfaces constant on ellipsoids, a property possessed in particular by the multivariate normal and t distributions. Some authors have discussed the estimation of the location and scale parameters for the elliptically contoured distributions, as well as, certain exact sampling distributional properties of the estimators of these parameters. But, as we show in the paper, the existing estimators for the scale and the kurtosis parameters of the elliptically contoured distributions are not consistent, which limits their uses in practice. As a remedial measure, we develop the consistent estimators for the location, scale, and kurtosis parameters of the elliptically contoured distributions, under a general cluster regression model.

Keywords and Phrases: Scale and kurtosis parameters; Maximum likelihood and moment estimators; Consistency; Cluster regression analysis.

AMS Classification: 62H12, 62H15.

1 Introduction

Suppose that p-dimensional random variables $Y_1, \ldots, Y_j, \ldots, Y_n$ are regarded as being distributed according to an np-dimensional elliptically contoured distribution (ECD). Under the assumption that elliptically contoured distributions, which constitute a generalization of the normal distribution, may be determined by the same parameters, many authors (Anderson et al, (1986)) have expressed the joint probability density function (p.d.f.) of $Y_1, \dots, Y_j, \dots, Y_n$ in the form

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

where μ is a p-dimensional location vector and $\Lambda > 0$ is a $p \times p$ scale matrix parameter. For the ECD, the covariance matrix Σ (say) is a scalar constant multiple of Λ. Thus, if n p-dimensional random vectors in (1) follow an elliptically contoured t distribution with ν (> 2) degrees of freedom (d.f.), then Σ is $c = \nu/(\nu - 2)$ multiple of Λ. In general, ν may be referred to as the shape parameter, which is unknown in practice. Note that as μ and Λ in (1) are free from the shape parameter, the ECDs determined by these parameters do not contain any information about the shape or kurtosis of the data. This observation motivates us (Sutradhar (1994)) to define the joint p.d.f. of $Y_1, \dots, Y_j, \dots, Y_n$ in the form

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

where $g_\kappa(\cdot)$ indicates that the density of the distribution depends on a kurtosis parameter κ which is a function of the shape parameter ν, as explained above. With regard to the estimation of the mean vector μ and the scale matrix Λ, Anderson et al (1986) maximized the likelihood function (1) and showed that the estimator of the mean vector is the estimator $\bar{Y} = \sum_{j=1}^{n} Y_j$ under normality and the estimator of the covariance matrix Σ is a constant multiple of the estimator $S = \sum_{j=1}^{n} (Y_j - \bar{Y})(Y_j - \bar{Y})'/n$ under normality. In the next section, we examine the consistency property of Anderson et al's estimators and show by a counter-example that their estimator of Σ is not consistent even if the ECD family or ν or κ is known. This leads to a potential problem for the inference about μ, as the covariance of $\bar{Y}$ in this case cannot be consistently estimated. In Section 3, we consider an ECD based clustered regression model and borrow the strengths from the independent clusters in the spirit of Liang and Zeger (1986), among others, to obtain consistent estimators for the regression, covariance matrix and the shape parameters of the model. Note that in this approach, the cluster size n can be as small as 1 but the number of clusters K (say) should be sufficiently large. We also simplify these estimators for the heteroscedastic univariate ($p = 1$) case with general n and K. This appears in Section 4. In this special case, the n univariate responses in a cluster are dependent on each other following an ECD structure, and similar to Liang and Zeger, one is mainly interested in the regression effects of covariates involved in the model.

2 Inconsistency of the Estimators of Covariance and Kurtosis Parameters

Recall that in our notations, for given ν or $\kappa(\nu)$, the maximum likelihood estimators of μ and Σ derived by Anderson et al (1986, Theorem 1, p. 56) are given by

$$\hat{\mu}_{\text{MLE}} = \bar{Y}, \;\; \hat{\Sigma}_{\text{MLE}} = (np/d^*)S \tag{3}$$

where d^* is a finite positive maximum of $d^{\frac{np}{2}} g^*_\kappa(d)$, with $g^*_\kappa(d)$ as the spherical joint density of $Y_1, \ldots, Y_j, \ldots, Y_n$. For example, under normality $g^*_\kappa(d) = (2\pi)^{-\frac{np}{2}} \exp(-d/2)$, and under t model with ν d.f., $g^*_\kappa(d) = \left[\nu^{\frac{\nu}{2}} \Gamma\{(\nu+np)/2\}/\pi^{\frac{np}{2}}\Gamma(\nu/2)\right] \{\nu + d\}^{-\frac{\nu+np}{2}}$. It is clear from (3) that $\hat{\mu}_{\text{MLE}} = \bar{Y}$ is a consistent estimator for μ. This is because, as $Y_1, \ldots, Y_j, \ldots, Y_n$ are uncorrelated, $\text{var}(\bar{Y}) = \Sigma/n$, which approaches zero as $n \to \infty$. In the same token, $\hat{\Sigma}_{\text{MLE}}$ in (3) will be consistent for Σ, provided the $p^2 \times p^2$ covariance matrix of $\hat{\Sigma}_{\text{MLE}}$ approaches to zero as $n \to \infty$. It does not, however, appear to be the case under ECDs. We show this inconsistency by using a counter-example based on the multivariate elliptic t contoured distribution. Consider following (2) that $Y_1, \ldots, Y_j, \ldots, Y_n$ have the np-dimensional elliptic t-distribution given by

$$C(\nu, n, p)|\Lambda|^{-\frac{n}{2}} \left\{\nu + \Sigma_{j=1}^n (y_j - \mu)'\Lambda^{-1}(y_j - \mu)\right\}^{-\frac{\nu+np}{2}}, \tag{4}$$

where $C(\nu, n, p) = \left[\nu^{\frac{\nu}{2}} \Gamma\{(\nu+np)/2\}/\pi^{\frac{np}{2}}\Gamma(\nu/2)\right]$, ν being the d.f. of the t-distribution. Although ν or $\kappa(\nu)$ is unknown in practice, for the time being we assume that ν is known. To simplify Anderson et al's (1986) estimators in (3), it is easy to compute d^* based on (4) which is np. Consequently, the consistency of $\hat{\Sigma}_{\text{MLE}}$ depends on the consistency of the normality based sample covariance matrix $S = \sum_{j=1}^{n} (Y_j - \bar{Y})(Y_j - \bar{Y})'/(n-1)$. Sutradhar and Ali (1989) have derived the exact sampling distribution of this sample covariance matrix under the general ECD (2). These authors have also simplified the distribution of S for the elliptic t distribution (4) as a special case and computed the variances and the pair-wise covariances of the elements of S matrix under the elliptic t distribution. Let S_{uv} be the (u, v)th element of the matrix S for $u, v = 1, \ldots, p$. Also, let $\Sigma^{\frac{1}{2}} = ((m_{h\ell}))$ be a symmetric matrix such that $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$ for $h, \ell = 1, \ldots, p$, where $\Sigma = \nu\Lambda/(\nu - 2)$. It then follows from (3.6) in Sutradhar and

Ali (1989) that the variances of the elements of S matrix are given by

$$\begin{aligned} \mathrm{var}(S_{uv}) &= (n-1)^{-2}(\nu-2)/(\nu-4)\left[(n-1)^2\left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 \right. \\ &\quad \left. + 2(n-1)\sum_{h=1}^{p} m_{uh}^2 m_{vh}^2 + (n-1)\sum_{h<\ell}(m_{uh}m_{v\ell}+m_{u\ell}m_{vh})^2\right] \\ &\quad - \left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 . \end{aligned} \tag{5}$$

It is clear from (5) that as $n \to \infty$, $\mathrm{var}(S_{uv})$ approaches to

$$\begin{aligned} Lt_{n\to\infty}\mathrm{var}(S_{uv}) &= \left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 \{(\nu-2)/(\nu-4)-1\} \\ &= \{2/(\nu-4)\}\left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 , \\ &= [2\nu^2/\{(\nu-2)^2(\nu-4)\}]\left(\sum_{h=1}^{p} \lambda_{uh}\lambda_{vh}\right)^2 , \end{aligned} \tag{6}$$

where λ_{uh} is the (u,h)th element of the $\Lambda^{\frac{1}{2}}$ matrix. Following the equation (3.5) in Sutradhar and Ali (1989), one may similarly compute the limiting pair-wise covariances of the elements of the S matrix. It is, however, clear that these limiting variances and covariances are free from n, and they approach to certain finite quantities based on the elements of the Λ matrix and the d.f. of the t-distribution. Consequently, the $\hat{\Sigma}_{\mathrm{MLE}}$ of Anderson et al (1986) is not consistent for Σ under the general elliptically contoured distribution (2). Further, as the sample covariance matrix S is not consistent for Σ, it then follows that the James-Stein and Stein's orthogonally invariant estimators (of Σ) constructed in Kubokawa and Srivastava (1997) are in fact the function of the elements of the inconsistent sample covariance matrix. The inconsistency of S (for Σ) also imply that there may not exist any consistent estimators for the shape parameter ν or kurtosis parameter $\kappa(\nu)$ under the ECD set-up (2). This we verify below under the elliptically contoured t distribution given in (4). To begin with, we, however, first note that there does not exist any maximum likelihood estimator for ν. This is because, for given μ and Λ, maximizing the joint p.d.f. (4) is equivalent to maximizing

$$C(\nu,n,p)|\Lambda|^{-\frac{n}{2}}\left\{1+\sum_{j=1}^{n}(y_j-\mu)'\Lambda^{-1}(y_j-\mu)/\nu\right\}^{-\frac{\nu+np}{2}} , \tag{7}$$

with respect to ν. For convenience, we consider $\mu = 0$ and $\Lambda = I$ without any loss of generality, and maximize the function

$$C(\nu, n, p)\left\{1 + \sum_{j=1}^{n} y_j' y_j/\nu\right\}^{-\frac{\nu+np}{2}}, \tag{8}$$

with respect to ν. Note that in (8) $C(\nu, n, p) = \left[\nu^{\frac{\nu}{2}}\Gamma\{(\nu + np)/2\}/\pi^{\frac{np}{2}}\Gamma(\nu/2)\right]$, which is an increasing function of ν for fixed n and p. Furthermore, as $\sum_{j=1}^{n} y_j' y_j$ and np are fixed for a given data set, it also follows that the spherical function in (8) is an increasing function of ν. Thus, it is intuitively clear that the joint density function in (8) is maximized at $\nu = \infty$. Consequently, there does not exist any maximum likelihood estimator of ν, as $\nu = \infty$ is the normal case. We have also verified the non-existence of the maximum likelihood estimator of ν numerically but do not include the numerical verification in the paper. Next, we show that even if a moment estimator for ν or $\kappa(\nu)$ exists, the estimator is not consistent under the general ECD set-up (2). To verify this, it is enough to show that the moment estimator $\hat{\nu}_M$ (say) is not consistent for ν for the elliptically contoured t distribution (4). Following Mardia (1970), a multivariate measure of kurtosis of the elliptical t-distribution may be written as

$$\beta_2 = \int [(y_j - \mu)'\Sigma^{-1}(y_j - \mu)]^2 dL, \tag{9}$$

where

$$L = C_1(\nu, n, p)|\Sigma|^{-\frac{n}{2}}\left\{(\nu - 2) + \sum_{j=1}^{n}(y_j - \mu)'\Sigma^{-1}(y_j - \mu)\right\}^{-\frac{\nu+np}{2}},$$

by (4). By direct integration, one may simplify (9) as

$$\beta_2 = \left(\frac{\nu - 2}{\nu - 4}\right) f(n, \sigma), \tag{10}$$

where

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

and $\sigma^{hh'}$ and $\sigma_{hh'}$ are the (h, h')th element of Σ^{-1} and Σ, respectively. It then follows that a moment estimator $\hat{\nu}_M$ of ν is given by

$$\hat{\nu}_M = 2[2\hat{\beta}_{2,M} - f(n, s)]/[\hat{\beta}_{2,M} - f(n, s)], \tag{11}$$

where $f(n, s)$ is computed by (10) after replacing Σ by the sample covariance matrix S, and $\hat{\beta}_{2,M}$ is the sample counter-part of β_2, given by

$$\hat{\beta}_{2,M} = \frac{1}{n}\sum_{j=1}^{n}[(y_j - \bar{y})' S^{-1}(y_j - \bar{y})]^2.$$

It is now easy to see that as S is not a consistent estimator for Σ, $\hat{\beta}_{2,M}$ and $f(n, s)$ in (11) cannot be consistent for their counterparts, β_2 and $f(n, \sigma)$, respectively. This leads to the conclusion that $\hat{\nu}_M$ in (11) is not a consistent estimator of ν. This, in turn, shows that the moment estimator of the shape parameter ν or the kurtosis parameter $\kappa(\nu)$ is not a consistent estimator under the general ECD set-up (2). In the next section, we consider a cluster regression set-up under the general ECD set-up (2) and construct suitable consistent estimators for all regression, covariance matrix and the shape or the kurtosis parameters of the ECD. In particular, the emphasis is given on the regression analysis with elliptically distributed errors.

3 ECD based Cluster Regression Model

Cluster correlated (or dependent) data constitute a set of independent multivariate responses $y_i = (y'_{i1}, \ldots, y'_{ij}, \ldots, y'_{in_i})'$, for $i = 1, \ldots, K$, together with an $n_i p \times pc$ matrix $X_i = (x'_{i1}, \ldots, x'_{ij}, \ldots, x'_{in_i})'$ where

$$x_{ij} = \begin{bmatrix} x'_{ij1} & 0' & \cdots & 0' \\ 0' & x'_{ij2} & \cdots & 0' \\ \vdots & \vdots & & \vdots \\ 0' & 0' & \cdots & x'_{ijp} \end{bmatrix}$$

is the $p \times pc$ covariate matrix for the jth $(j = 1, \ldots, n_i)$ individual in the ith cluster, and $0'$ is the $1 \times c$ null vector. In this set-up, a p-dimensional response vector y_{ij} and the $p \times pc$ matrix of covariates are observed for the jth individual in the ith cluster / family / group. The data of this type may be modelled as

$$y_i = X_i\beta + \epsilon_i, \tag{12}$$

where $\beta = (\beta_1, \ldots, \beta_h, \ldots, \beta_p)'$ with $\beta_h = (\beta_{h1}, \ldots, \beta_{hc})'$ and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ij}, \ldots, \epsilon_{in})'$ with $\epsilon_{ij} = (\epsilon_{ij1}, \ldots, \epsilon_{ijk}, \ldots, \epsilon ijp)'$. As opposed to Liang and Zeger's (1986) generalized linear model based cluster regression problem, the present model is linear but with a special error structure having general elliptically contoured distribution (2) for ϵ_i. More specifically, it is assumed that the $n_i p$ observations in y_i, follow the ECD given by

$$|\Lambda|^{-\frac{n_i}{2}} g_\kappa((y_i - X_i\beta)'(I_{n_i} \otimes \Lambda^{-1})(y_i - X_i\beta)), \tag{13}$$

where I_{n_i} is the $n_i \times n_i$ identity matrix. Note that because of the form $g_\kappa(\cdot)$, the p-dimensional observations $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ are pairwise uncorrelated but not necessarily independent. Further note that the ECD in (13) also may be written as

$$|\Lambda|^{-\frac{n_i}{2}} g_\kappa\left(\sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)'\Lambda^{-1}(y_{ij} - x_{ij}\beta)\right). \tag{14}$$

The main purpose of this section is to obtain a consistent estimator $\hat{\beta}_{\text{GLS}}$ (say) of β, as well as to estimate the covariance matrix of $\hat{\beta}_{\text{GLS}}$, consistently. Under the present set-up, the well known generalized least square equations for β may be written as

$$\sum_{i=1}^{K} X_i^T(I_{n_i} \otimes \hat{\Sigma}^{-1})(y_i - X_i\beta) = 0, \tag{15}$$

yielding

$$\hat{\beta}_{\text{GLS}} = \left[\sum_{i=1}^{K} X_i^T(I_{n_i} \otimes \hat{\Sigma}^{-1})X_i\right]^{-1} \sum_{i=1}^{K} X_i^T(I_{n_i} \otimes \hat{\Sigma}^{-1})y_i. \tag{16}$$

As (15) is an unbiased estimating equation, it then follows under some mild regularity conditions on X_i that $\hat{\beta}_{\text{GLS}}$ is a consistent estimator for β and its covariance matrix can be consistently estimated by

$$\hat{V}(\hat{\beta}_{\text{GLS}}) = \left[\sum_{i=1}^{K} X_i^T(I_{n_i} \otimes \hat{\Sigma}^{-1})X_i\right]^{-1}, \tag{17}$$

provided $\hat{\Sigma}$ is a consistent estimator for Σ.

3.1 Consistent estimate of Σ

3.1.1 A pooled estimate

For known β, define a sample covariance matrix $S_i = \sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)(y_{ij} - x_{ij}\beta)'/n_i$ based on n_i p-dimensional observation under the ith cluster. As for simplicity, it was shown in the last section in the context of elliptic t distribution that this S_i matrix is not a consistent estimator for Σ, we now pool the information from K independent clusters and estimate Σ as

$$\hat{\Sigma} = \frac{1}{K}\sum_{i=1}^{K} S_i, \tag{18}$$

and examine its consistency under the general ECD set-up (2). Since the $p^2 \times p^2$ covariance matrix of S_i does not depend on any cluster, we compute the covariance matrix for S_1 without any loss of generality. For $u, v, h, \ell = 1, \ldots, p$, let $S_{1,uv}$ and $S_{1,h\ell}$ be two general elements of the $S_1 = \sum_{j=1}^{n_1}(y_{1j} - x_{1j}\beta)(y_{1j} - x_{1j}\beta)'/n_1$ matrix. More specifically,

$$S_{1,uv} = \sum_{j=1}^{n_1}(y_{1ju} - x'_{1ju}\beta_u)(y_{1jv} - x'_{1jv}\beta_v)/n_1,$$

and

$$S_{2,h\ell} = \sum_{j=1}^{n_1}(y_{1jh} - x'_{1jh}\beta_h)(y_{1j\ell} - x'_{1j\ell}\beta_\ell)/n_1. \tag{19}$$

Recall from (2) that κ is the kurtosis parameter for every component of Y_{ij}. For example, for multivariate elliptic t distribution, $\kappa = 2/(\nu - 4)$, where ν is the shape parameter or degrees of freedom of the t-distribution. Note that this ν or κ is unknown too in practice. It then follows from (19) [Muirhead (1982, p. 40-49)] that

$$\begin{aligned}\mathrm{cov}(S_{1,uv}, S_{1,h\ell}) &= \frac{1}{n_1^2}[n_1(\kappa+1)(\sigma_{uv}\sigma_{h\ell} + \sigma_{uv}\sigma_{v\ell} + \sigma_{u\ell}\sigma_{vh}) \\ &\quad + n_1(n_1-1)(\kappa+1)\sigma_{uv}\sigma_{h\ell}] - \sigma_{uv}\sigma_{h\ell},\end{aligned}$$

which may be simplified as

$$\mathrm{cov}(S_{1,uv}, S_{1,h\ell}) = \kappa\sigma_{uv}\sigma_{h\ell} + (\kappa+1)(\sigma_{uh}\sigma_{v\ell} + \sigma_{u\ell}\sigma_{vh})/n_1, \tag{20}$$

where σ_{uv}, for example, is the (u, v)th element of the Σ matrix. Note that even if the cluster size n_i approaches to ∞, this covariance in (20) does not reduce to zero. But, as $S_1, \ldots, S_i, \ldots, S_K$ are sample covariance matrices from independent clusters, it is clear from (18) that the $p^2 \times p^2$ covariance matrix of $\hat{\Sigma}$ approaches zero as number of clusters K approaches ∞. This covariance matrix estimate, therefore, may be used in (16) and (17) to estimate β and its confidence intervals. Note that in the present approach β and Σ are estimated iteratively by using (16) and (18).

3.1.2 Independent clusters based direct estimate for the case $n_i = n$

Unlike in the last section, this approach requires $n_i = n$ for all clusters $i = 1, \ldots, K$. For $n_i = n$, $y_1, \ldots, y_i, \ldots, y_K$ are independently distributed $np \times 1$ random elliptic vectors with mean and covariance matrix given by

$$E(Y_i) = X_i\beta, \text{ and } \mathrm{cov}(Y_i) = I_n \otimes \Sigma$$

respectively, for all $i = 1, \ldots, K$. Further it is assumed that every component of Y_i has the same kurtosis κ. Next, let

$$S^*(K - pc) = \sum_{i=1}^{K}(Y_i - X_i\hat{\beta}_{\mathrm{GLS}})(Y_i - X_i\hat{\beta}_{\mathrm{GLS}})'/(K - pc)$$

be the sample covariance matrix constructed based on K np-dimensional independent vectors $Y_1, \ldots, Y_i, \ldots, Y_K$. It then follows that (Muirhead (1982, p. 42)) the asymptotic $(K \to \infty)$ distribution of

$$Z(K - pc) = (K - pc)^{\frac{1}{2}}[S^*(K - pc) - I_n \otimes \Sigma] \tag{21}$$

is normal with mean zero. The covariance, expressed in terms of the kurtosis and elements of $\Sigma^* = I_n \otimes \Sigma = (\sigma^*_{uv})$ are given by

$$\begin{aligned} \mathrm{cov}(Z_{uv}, Z_{h\ell}) &= \kappa[(\sigma^*_{uv}\sigma^*_{h\ell} + \sigma^*_{uh}\sigma^*_{v\ell} + \sigma^*_{u\ell}\sigma^*_{vh})] \\ &\quad +(\sigma^*_{uh}\sigma^*_{v\ell} + \sigma^*_{u\ell}\sigma^*_{vh}). \end{aligned} \tag{22}$$

It then follows from (21) that

$$\begin{aligned} \mathrm{cov}(S^*_{uv}, S^*_{h\ell}) &= (K - pc)^{-1}[\kappa(\sigma^*_{uv}\sigma^*_{h\ell} + \sigma^*_{uh}\sigma^*_{v\ell} + \sigma^*_{u\ell}\sigma^*_{vh}) \\ &\quad +(\sigma^*_{uh}\sigma^*_{v\ell} + \sigma^*_{u\ell}\sigma^*_{vh})], \end{aligned} \tag{23}$$

showing that S^* is a consistent estimator for $I_n \otimes \Sigma$. Note that for the elliptic t distribution $\kappa = 2/(\nu - 4)$ and for normal distribution $\kappa = 0$. One of the drawbacks of the 'independent cluster based approach' to estimate Σ is that all the off-diagonal null matrices in $I_n \otimes \Sigma$ are estimated by corresponding off-diagonal non-null matrices of S^*. Furthermore, Σ matrix itself are estimated by n different diagonal matrices of S^*, leading to non-unique estimates for the same Σ. Consequently, the GLS estimate of β computed by

$$\hat{\beta}^*_{\mathrm{GLS}} = \left[\sum_{i=1}^{K} X_i^T {S^*}^{-1} X_i\right]^{-1} \sum_{i=1}^{K} X_i^T {S^*}^{-1} y_i \tag{24}$$

will be less efficient as compared to $\hat{\beta}_{\mathrm{GLS}}$ computed by (16). This problem of non-unique estimates for Σ does not, however, arise for $n = 1$.

3.2 Consistent estimate for the kurtosis parameter

Note that the inferences about β do not require the knowledge of the kurtosis parameter of the ECD, rather they require the consistent estimate of Σ, which we have already discussed in the previous sections. It, however, becomes a different situation when one would like to infer about the covariance matrix Σ. This is clear from (20) and (23) where it was shown that the covariance matrix of $\hat{\Sigma}$ is a function of the kurtosis parameter of the general ECD (2). In order to estimate the kurtosis parameter κ, we follow Mardia (1970) and define a measure of kurtosis of ECD given by

$$\beta_{2i} = E[Y_i - X_i\beta)'(I_{n_i} \otimes \Sigma^{-1})(Y_i - X\beta)]^2. \tag{25}$$

Note that the same measure β_2 was used in Section 2 in order to examine the consistency of the estimator of the shape parameter ν of the elliptic t-distribution. It was shown there that $\hat{\nu}_M$ is not consistent for ν. We are, however, dealing with a slightly different situation here, as in the present set-up, we have n_ip-dimensional elliptic observations from the ith cluster and we have altogether K independent clusters to exploit to obtain a consistent estimator of the shape parameter or kurtosis parameter. Writing $\left(I_{n_i} \otimes \Sigma^{-\frac{1}{2}}\right)(Y_i - X\beta) = Z_i$, one can compute the expectation in (25) as

$$\begin{aligned} \beta_{2i} &= E[Z_i'Z_i]^2 \\ &= E\left[\sum_{h=1}^{n_ip} Z_{ih}^2\right]^2 \\ &= E\left[\sum_{h=1}^{n_ip} Z_{ih}^4 + \sum_{h\neq\ell}^{n_ip} Z_{ih}^2 Z_{i\ell}^2\right], \end{aligned} \tag{26}$$

where Z_i has a n_ip-dimensional unit spherical distribution. As we have assumed that each component of Y_i has the same kurtosis parameter κ, it then follows that (cf. Muirhead (1982, p. 41))

$$E(Z_{ih}^4) = 3(\kappa + 1) \text{ and } E(Z_{ih}^2) = 1,$$

yielding β_{2i} as

$$\begin{aligned} \beta_{2i} &= 3n_ip(\kappa + 1) + n_ip(n_ip - 1) \\ &= n_ip\{3\kappa + n_ip + 2\}. \end{aligned} \tag{27}$$

Next, as $y_1, \ldots, y_i, \ldots, y_K$ are independent observations each having an ECD, it is easy to see that

$$\hat{\beta}_2 = \frac{1}{K}\sum_{i=1}^{K}[(y_i - X_i\hat{\beta}_{\text{GLS}})'(I_{n_i} \otimes \hat{\Sigma}^{-1})(y_i - X_i\hat{\beta}_{\text{GLS}})]^2$$

approaches in probability to

$$3p\kappa\sum_{i=1}^{K} n_i/K + p^2\sum_{i=1}^{K} n_i^2/K + 2p\sum_{i=1}^{K} n_i/K.$$

One may then obtain a consistent estimator for the kurtosis parameter given by

$$\hat{\kappa} = \frac{K}{3p\sum_{i=1}^{K} n_i}\left[\hat{\beta}_2 - p^2\sum_{i=1}^{K} n_i^2/K - 2p\sum_{i=1}^{K} n_i/K\right]. \tag{28}$$

4 Univariate Heteroscedastic Models: A Special Case

4.1 Single cluster based inference

Suppose that the ith cluster contains n_i individuals and each of the individuals provides a univariate response. Let y_{ij} be the response of the jth $(j = 1, \ldots, n_i)$ individual in the ith cluster, and $x_{ij} = (x_{ij1}, \ldots, x_{ijc})'$ be a $c \times 1$ vector of covariates associated with y_{ij}. Further suppose that the n_i responses in the ith cluster follow an n_i-dimensional elliptically contoured distribution (2) with

$$E(Y_i) = X_i\beta,\ \text{var}(Y_i) = \Sigma = \ \text{diag}(\sigma_1^2, \ldots, \sigma_{n_i}^2) \tag{29}$$

and each component of y_i has the same kurtosis parameter κ. Here X_i is the $n_i \times c$ matrix of covariates and β is the $c \times 1$ regression vector. This is a linear model case but the joint p.d.f. of $y_{i1}, \ldots, y_{in_i}$ is given by

$$C(\kappa)\Pi_{j=1}^{n_i}(\sigma_j^2)^{-\frac{1}{2}} g_\kappa\left(\sum_{j=1}^{n_i}\{(y_{ij} - x_{ij}'\beta)^2/\sigma_j^2\}\right). \tag{30}$$

Inferences about β require the estimation of $\sigma_1^2, \ldots, \sigma_{n_i}^2$. Note, however, that one cannot estimate these variance components from a single cluster, namely, from the ith cluster. This is because, as the cluster size gets larger, the number of variance parameters increases too. Much more serious situation than this is that even if $\sigma_1^2 = \ldots, \sigma_{n_i}^2 = \sigma^2$, say, unlike in the normal case, this single variance component σ^2 cannot be consistently estimated by the sample variance defined based on the n_i elements of the ith cluster. To verify this, for known β, define the sample variance from the ith cluster as

$$s_i^2 = n_i^{-1}\sum_{j=1}^{n_i}(y_{ij} - x_{ij}'\beta)^2. \tag{31}$$

Now, as it was discussed in Sections 2 and 3, this estimator has the variance given by (20) as

$$v(s_i^2) = \kappa\sigma^4 + (\kappa + 1)(2\sigma^4/n_i)$$

under general ECD (1.2) and the variance

$$v(s_i^2) = \{2/(\nu-4)\}\sigma^4 + \{(\nu-2)/(\nu-4)\}2\sigma^4/n_i$$

under the elliptic t distribution with ν degrees of freedom. Thus, it is clear that even if n_i is sufficiently large, σ^2 cannot be estimated consistently. This, in turn, affects the inference about β, in particular, the construction of its confidence interval.

4.2 Inference from combined clusters

As discussed in Section 3.1, the above problem of estimation of σ^2 can, however, be removed by pooling the information from all K independent clusters. Thus

$$\hat{\sigma}^2 = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - x'_{ij}\beta)^2/n_i \tag{32}$$

is a consistent estimate of σ^2 (as $K \to \infty$). Note, however, that the exact sampling distribution theory for $\hat{\sigma}^2 = \sum_{i=1}^{K} s_i^2/K$ is extremely complicated under general ECD. If β is unknown, which is usually the case, it is replaced by $\hat{\beta} = \left(\sum_{i=1}^{K} X'_i X_i\right)^{-1} \sum_{i=1}^{K} X'_i y_i$ and $\hat{\sigma}^2$ in (32) reduces to

$$\hat{\sigma}^2 = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - x'_{ij}\hat{\beta})^2/(n_i - c). \tag{33}$$

It may be remarked that in the cluster regression set-up, even if the variance components are different, for the special case $n_i = n$, they can be consistently estimated, in the spirit of Liang and Zeger (1986), by using

$$\hat{\sigma}_j^2 = \frac{1}{K}\sum_{i=1}^{K}(y_{ij} - x'_{ij}\hat{\beta})^2, \tag{34}$$

for all $j = 1, \ldots, n$, as K clusters then work as K independent repeated samples.

Acknowledgements

This research was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author would like to thank the referee for suggestions that improved the paper.

References

[1] 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.

[2] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, 1, 361-379, University of California Press, Berkeley.

[3] KUBOKAWA, T. and SRIVASTAVA, M. S. (1997). Robust improvements in estimation of mean and covariance matrices in elliptically contoured distribution. *Technical Report No. 9714*, Department of Statistics, University of Toronto.

[4] LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

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

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

[7] SUTRADHAR, B. C. (1994). On cluster regression and factor analysis models with elliptic t errors. In *IMS Lecture Notes- Monograph Series,* eds. T. W. Anderson, K. T. Fang and I. Olkin, 24, 370-383.

[8] SUTRADHAR, B. C. and ALI, M. M. (1986). Estimation of parameters of a regression model with a multivariate t error variable. *Commun. Statist. Theory and Methods*, 15, 429-450.

[9] 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.

BLANK PAGE