REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

AD-A271 031

Public r... gatheri... collectic... Davis Hi... 1 hour per response, including the time for reviewing instructions, searching existing data sources, ...ollection of information. Send comments regarding this burden estimate or any other aspect of this ...ington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...gement and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AG	3. REPORT TYPE AND DATES COVERED Reprint
4. TITI Title shown on Reprint	5. FUNDING NUMBERS (2) DAAH04-93-G-0030
6. AUTHOR(S) Authors listed on Reprint	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Pennsylvania State Univ. University Park, PA 26802	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARO 30529.25-MA-SDI

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

ABSTRACT SHOWN ON REPRINT

93-24883

14. SUBJECT TERMS			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Journal of Statistical Planning and Inference 36 (1993) 77-90
North-Holland

MANOVA type tests under a convex discrepancy function for the standard multivariate linear model

Z.D. Bai

Temple University, Philadelphia, PA, USA

C. Radhakrishna Rao and L.C. Zhao*

Pennsylvania State University, University Park, PA, USA

Received 8 November 1990; revised manuscript received 5 February 1992
Recommended by P.K. Sen

Abstract: We provide the M-theory for the standard multivariate linear model $Y = XB + E$, where Y is $n \times p$ matrix of observations, X is $n \times m$ design matrix, B is $m \times p$ matrix of unknown parameters and E is $n \times p$ matrix of errors with the row vectors independently distributed. Two test criteria based on the roots of determinantal equations are proposed for testing linear hypotheses of the form $P'B = C_0$, where P is a matrix of rank q. The tests are similar to those considered in MANOVA using least squares techniques. One of them is the Wald type statistic and another is the Rao's score type statistic. The asymptotic distributions of these test statistics are derived. Consistent estimates of nuisance parameters are obtained for use in computing the test statistics.

The M-method of estimation considered is the minimization of $\sum\rho(e_i)$, where ρ is a convex function and e_i is the i-th row vector in $(Y-XB)$. All results are derived under a minimal set of conditions.

AMS Subject Classification: 62H15, 62H10.

Key words and phrases: MANOVA; M-estimation; Rao's score test; roots of determinantal equation; Wald test.

1. Introduction

In a recent paper Bai, Rao and Wu (1992) considered the problem of estimation and testing under the M-theory for the model

$$Y_i = X_i'\beta + \varepsilon_i, \tag{1.1}$$

Correspondence to: Prof. C.R. Rao, Dept. of Statistics, Centre for Multivariate Analysis, The Pennsylvania State University, 417 Classroom Building, University Park, PA 16802, USA.

Research sponsored by the Air Force Office of Scientific Research under Grant AFOSR-89-0279 and the U.S. Army Research Office under Grant DAAL03-89-K-0139.

* On leave from University of Science and Technology of China.

where Y_i is a p-vector of observations, ε_i is a p-vector of errors, $\{X_i\}$ is a design sequence of $m \times p$ matrices and β is an m-vector of unknown parameters. The discussion was confined to estimation of β by minimizing

$$\sum_{i=1}^{n} \varrho(Y_i - X_i'\beta) \tag{1.2}$$

choosing any convex function ϱ. The asymptotic distribution of $\hat{\beta}$, the estimate so obtained, was derived. For testing the hypothesis $P'\beta = C_0$, the test criterion proposed was the likelihood ratio type

$$\min_{P'\beta = C_0} \sum \varrho(Y_i - X_i'\beta) - \min_{\beta} \sum \varrho(Y_i - X_i'\beta), \tag{1.3}$$

which, under suitable normalization, has an asymptotic distribution which is a mixture of chi-squares.

We now consider a special case of (1.1), the standard multivariate linear model

$$Y_i = B'X_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.4}$$

where Y_i and ε_i are as in the model (1.1), B is an $m \times p$ matrix of regression coefficients and $\{X_i\}$ is a design sequence of m-vectors. As in (1.2), we estimate B by minimizing

$$\sum_{i=1}^{n} \varrho(Y_i - B'X_i), \tag{1.5}$$

where ϱ is a covex function, and develop MANOVA type analysis leading to test criteria based on the roots of a determinantal equation for testing hypotheses of the type $P'B = C_0$, where P is $m \times q$ matrix of rank q.

2. Notations and assumptions

Let $\psi(u)$ be a choice of a subgradient of ϱ at $u = (u_1, \ldots, u_p)'$. [A p-vector $\psi(u)$ is said to be a subgradient of ϱ at u, if $\varrho(z) \geqslant \varrho(u) + (z-u)'\psi(u)$ $\forall z \in R^p$.] Note that if ϱ is differentiable at u according to the usual definition, ϱ has a unique subgradient at u and vice-versa. In this case

$$\psi(u) = \nabla \varrho(u) \triangleq \left(\frac{\partial \varrho}{\partial u_1}, \ldots, \frac{\partial \varrho}{\partial u_p} \right)^1.$$

Denote by $\mathscr{D}$ the set of points where ϱ is not differentiable. This is, in fact, the set of points where ψ is discontinuous, which is the same for all choices of ψ. It is well-known that $\mathscr{D}$ is topologically a F_σ set of Lebesgue measure zero (ref. Rockafeller (1970), p. 218 and Section 25).

We assume that $\psi(u)$ is measurable and make the following assumptions as in Bai, Rao and Wu (1992):

(A_1) The common distribution function F of ε_i satisfies $F(\mathscr{D})=0$. (This ensures that certain functionals of ψ which appear in our discussion have unique values.)
(A_2) $E\psi(\varepsilon_1+u)=\Lambda u+o(\|u\|)$ as $\|u\|\to 0$, where $\Lambda>0$ is a $p\times p$ constant matrix.
(A_3) $E\|\psi(\varepsilon_1+u)\|^2$ is finite for small $\|u\|$ and is continuous at $u=0$ as a function of u.

(A_4) $$E[\psi(\varepsilon_1)][\psi(\varepsilon_1)]'=\Gamma>0.$$

(A_5) $$S_n=\sum_{i=1}^n X_iX_i'>0,$$

and

$$d_n^2=\max_{1\leqslant i\leqslant n} X_i'S_n^{-1}X_i\to 0 \quad \text{as } n\to\infty.$$

We denote by $\hat{B}$ and $\tilde{B}$ any values of B which minimize

$$\sum_{i=1}^n \varrho(Y_i-B'X_i)$$

respectively without any restriction and subject to the restriction

$$P'B=C_0 \tag{2.1}$$

specified as a hypothesis, where P is a $m\times q$ matrix of rank q. Further let

$$\xi(B)=\sum_{i=1}^n X_i[\psi(Y_i-B'X_i)]' \tag{2.2}$$

which is an $m\times p$ matrix.

For testing the hypothesis $P'B=C_0$, we propose two alternative test criteria. One is based on the roots of the determinantal equation

$$|W_n-\theta\hat{\Lambda}^{-1}\hat{\Gamma}\hat{\Lambda}^{-1}|=0, \tag{2.3}$$

where

$$W_n=(P'\hat{B}-C_0)'(P'S_n^{-1}P)^{-1}(P'\hat{B}-C_0) \tag{2.4}$$

is the Wald type statistic, and $(\hat{\Lambda},\hat{\Gamma})$ is a consistent estimate of (Λ,Γ), the matrix parameters defined in assumptions (A_2) and (A_4) respectively. In Section 5 of this paper, we discuss the estimation of (Λ,Γ). Another test is based on the roots of the determinantal equation

$$|R_n-\theta\hat{\Gamma}|=0, \tag{2.5}$$

where

$$R_n=\xi(\tilde{B})'S_n^{-1}\xi(\tilde{B}) \tag{2.6}$$

is the Rao's score type statistic (see Rao (1948)), and $\hat{\Gamma}$ is a consistent estimate of Γ. The asymptotic distribution of the roots of (2.3) or (2.5) is the same as that in the normal theory, and hence the tests proposed by Fisher and Hsu (see for instance Rao (1973, pp. 556–560)) can be used.

It may be noted that tests of the above type were suggested by Sen (1982) and

Singer and Sen (1985) in the multivariate situation under methods of M-estimation and assumptions different from ours, and by Schrader and Hettmansperger (1980) in the univariate case. Some papers of related interest are by Inagaki (1973), Heiler and Willers (1988) and Jurečková (1983). It may be seen that our conditions are somewhat simpler in view of the convexity of the loss function.

In Section 3, we state the main theorems and in Section 4, we provide proofs under what we believe to be a minimal set of conditions. A new feature of the paper is the discussion on consistent estimation of the nuisance parameters Λ and Γ without making any further assumptions on ψ.

The results of the paper could be extended to other methods of M-estimation such as those with scale invariance or those based on estimating equations only. But they seem to need heavy assumptions for a rigorous treatment. It would also be of some interest to consider rates of convergence and related problems. We hope to consider such problems in future research.

3. The main theorems

For convenience, we write

$$X_{in}=S_n^{-1/2}X_i,\qquad P_n'=(P'S_n^{-1}P)^{-1/2}P'S_n^{-1/2},\tag{3.1}$$

so that

$$\sum_{i=1}^{n}X_{in}X_{in}'=I_m,\qquad P_n'P_n=I_q,\tag{3.2}$$

$$U_n'=\Lambda^{-1}\sum_{i=1}^{n}\psi(\varepsilon_i)X_{in}'P_n=(u_{1n},\dots,u_{qn}),\tag{3.3}$$

$$V_n'=\sum_{i=1}^{n}\psi(\varepsilon_i)X_{in}'P_n=(v_{1n},\dots,v_{qn}).\tag{3.4}$$

We also consider a sequence of alternatives to the null hypothesis $P'B=C_0$

$$\mathrm{H}_n\colon P'(B-B_0)=P'\Delta_n,\tag{3.5}$$

where B_0 and Δ_n are known $m\times p$ matrices such that

$$P'B_0=C_0\quad\text{and}\quad \|S_n^{1/2}\Delta_n\|=\mathrm{O}(1),\tag{3.6}$$

and denote

$$\Theta_n=P_n'S_n^{1/2}\Delta_n=(P'S_n^{-1}P)^{-1/2}P'\Delta_n.\tag{3.7}$$

It is easily seen that $u_{1n},\dots,u_{qn}$ are asymptotically independent with the common limiting distribution $\mathrm{N}_p(0,\Lambda^{-1}\Gamma\Lambda^{-1})$, so that the limiting distribution of $U_n'U_n$ is central Wishart on q degrees of freedom, $W_p(q,\Lambda^{-1}\Gamma\Lambda^{-1})$. Similarly $v_{1n},\dots,v_{qn}$ are asymptotically independent with the common limiting distribution $\mathrm{N}(0,\Gamma)$, so that the limiting distribution of $V_n'V_n$ is central Wishart on q degrees of freedom, $W_p(q,\Gamma)$.

We have the following theorems concerning the asymptotic distributions of W_n and R_n under the null hypothesis and also under the sequence of alternative hypotheses (3.5).

Theorem 3.1. *Assume that under the model* (1.4), *the assumptions* (A_1)-(A_5) *and condition* (3.6) *on the sequence of alternative hypotheses hold. Then*

$$W_n = (U_n + \Theta_n)'(U_n + \Theta_n) + o_p(1) \quad \text{as } n \to \infty. \tag{3.8}$$

Especially, if the null hypothesis H_0 *holds or* $\|S_n^{1/2}\Delta_n\| \to 0$ *as* $n \to \infty$, *then the asymptotic distribution of* W_n *is the central Wishart*, $W_p(q, \Lambda^{-1}\Gamma\Lambda^{-1})$. *If the local alternatives* $\Theta_n = (P'S_n^{-1}P)^{-1/2}P'\Delta_n$ *has a limit* $\Theta \neq 0$ *as* $n \to \infty$, *then the asymptotic distribution of* W_n *is the noncentral Wishart*, $W_p(q, \Lambda^{-1}\Gamma\Lambda^{-1}, \Theta'\Theta)$. [*See Rao* (1973, *p.* 534) *for the definition of noncentral Wishart distribution.*]

Theorem 3.2. *Suppose that under the model* (1.4), *the assumptions* (A_1)-(A_5) *are satisfied, and condition* (3.6) *holds. Then*

$$R_n = (V_n + \Theta_n\Lambda)'(V_n + \Theta_n\Lambda) + o_p(1) \quad \text{as } n \to \infty. \tag{3.9}$$

Especially, if H_0 *holds or* $\|S_n^{1/2}\Delta_n\| \to 0$ *as* $n \to \infty$, *the asymptotic distribution of* R_n *is the central Wishart*, $W_p(q,\Gamma)$. *If the local alternatives* Θ_n *has the limit* $\Theta \neq 0$, *then the asymptotic distribution of* R_n *is the noncentral Wishart*, $W_p(q, \Gamma, \Lambda\Theta'\Theta\Lambda)$.

Note: The test based on W_n involves two nuisance matrix parameters Γ and Λ, both of which need to be estimated for computing the test criteria. On the other hand, the test based on R_n involves only the nuisance matrix parameter Γ, which only needs to be estimated. Further, the local power for the sequence of alternatives considered depends on the magnitude of the roots of the equation

$$|\Theta'\Theta - \lambda\Lambda^{-1}\Gamma\Lambda^{-1}| = 0 \tag{3.10}$$

for the test based on W_n and on the roots of the equation

$$|\lambda\Theta'\Theta\Lambda - \lambda\Gamma| = 0 \tag{3.11}$$

for the test based on R_n. Since the roots of (3.10) and (3.11) are the same, the two alternative tests are equally efficient asymptotically. In such a case, the test based on R_n may be preferred to that based on W_n as only Γ has to be estimated in the former case and both Γ and Λ have to be estimated in the latter case. This statement is true only for large samples. The relative merits of these two tests remain to be investigated in small samples.

4. Proof of the main theorems

In the following, for a set A, $I(A)$ denotes its indicator function. We write

$$B^* = \operatorname{vec} B = \operatorname{vec}(\beta_1 \vdots \cdots \vdots \beta_p) = (\beta_1', \ldots, \beta_p')', \tag{4.1}$$

and the same notation applies to other matrices.

To prove the theorems stated in Section 3, we need some lemmas. Without loss of generality, we assume that $C_0 = 0$ in (2.1), i.e., $B_0 = 0$ in (3.5). There exists an $m \times (m-q)$ matrix K of rank $m-q$ such that

$$P'K = 0. \tag{4.2}$$

Without loss of generality we can assume that $K'\Delta_n = 0$. The hypotheses H_0 and H_n can be written as

$$\mathrm{H}_0\colon\ B = KM_0 \quad \text{for some } (m-q)\times p \text{ matrix } M_0,$$

$$\mathrm{H}_n\colon\ B = KM_0 + \Delta_n.$$

Define

$$P_n' = (P'S_n^{-1}P)^{-1/2}P'S_n^{-1/2}, \qquad K_n = S_n^{1/2}K(K'S_nK)^{-1/2}. \tag{4.3}$$

Then

$$K_n'K_n = I_{m-q}, \qquad P_n'P_n = I_q, \qquad P_n'K_n = 0. \tag{4.4}$$

If H_n holds, the model (1.4) can be rewritten as

$$Y_i = B_n'X_{in} + \varepsilon_i, \tag{4.5}$$

where $X_{in} = S_n^{-1/2}X_i$, as defined in (3.1),

$$B_n = K_nM_n + P_n\Theta_n, \tag{4.6}$$

$$M_n = (K'S_nK)^{1/2}M_0 + K_n'S_n^{1/2}\Delta_n, \tag{4.7}$$

and Θ_n is defined by (3.7).

Put $M_{0n} = (K'S_nK)^{1/2}M_0$. The model (1.4) under H_0 has the form

$$Y_i = (K_nM_{0n})'X_{in} + \varepsilon_i. \tag{4.8}$$

Denote by $\hat{M}_n$ the M-estimate of M_{0n}, i.e., $\hat{M}_n$ is such that

$$\sum_{i=1}^n \varrho(Y_i - \hat{M}_n'K_n'X_{in}) = \min_{M:(m-q)\times p} \sum_{i=1}^n \varrho(Y_i - M'K_n'X_{in}). \tag{4.9}$$

Note that the restricted M-estimate of B_n is $\tilde{B}_n = S_n^{1/2}\tilde{B} = K_n\hat{M}_n$.

Lemma 4.1. *Suppose that* (A_1)-(A_5), (3.5) *and* (3.6) *are satisfied. For any constant* $c>0$ *we have*

$$\sup_{\|G-M_n\|\leqslant c} \left\| \sum_{i=1}^n \{\psi(Y_i - G'K_n'X_{in}) - \psi(Y_i - B_n'X_{in})\} \otimes X_{in}\right.$$

$$+(\Lambda\otimes K_n)(G^*-M_n^*)-(\Lambda\otimes P_n)\Theta_n^*\Big\| \to 0 \quad \textit{in pr.}, \tag{4.10}$$

and

$$\sup_{\|G-M_n\|\leqslant c}\Big|\sum_{i=1}^{n}\{\varrho(Y_i-G'K_n'X_{in})-\varrho(Y_i-B_n'X_{in})\}$$
$$+f_n(\Theta_n^*)+g_n(G^*-M_n^*)\Big| \to 0 \quad \textit{in pr.}, \tag{4.11}$$

where

$$f_n(\Theta_n^*)=\Theta_n^{*\prime}\sum_i \psi(\varepsilon_i)\otimes(P_n'X_{in})-\tfrac{1}{2}\Theta_n^{*\prime}(\Lambda\otimes I_q)\Theta_n^*, \tag{4.12}$$

$$g_n(G^*-M_n^*)=(G^*-M_n^*)'\sum_i \psi(\varepsilon_i)\otimes(K_n'X_{in})$$
$$-\tfrac{1}{2}(G^*-M_n^*)'(\Lambda\otimes I_{m-q})(G^*-M_n^*). \tag{4.13}$$

Note that (4.10) and (4.11) can be simply rewritten as

$$\sup_{\|B\|\leqslant c}\Big\|\sum_{i=1}^{n}\{\psi(\varepsilon_i-B'X_{in})-\psi(\varepsilon_i)\}\otimes X_{in}+(\Lambda\otimes I_m)B^*\Big\| \to 0 \quad \text{in pr.} \tag{4.10$'$}$$

and

$$\sup_{\|B\|\leqslant c}\Big|\sum_{i=1}^{n}\{\varrho(\varepsilon_i-B'X_{in})-\varrho(\varepsilon_i)+B^{*\prime}(\psi(\varepsilon_i)\otimes X_{in})\}$$
$$-\tfrac{1}{2}B^{*\prime}(\Lambda\otimes I_m)B^*\Big| \to 0 \quad \text{in pr.} \tag{4.11$'$}$$

The proofs of (4.11)′ and (4.10)′ are similar to those of Theorems 2.1 and 2.3 in Bai, Rao and Wu (1992). Note that when we use Theorem 25.7 in Rockafellar (1970), we could remove the differentiable condition on $\{f_i\}$, regard $\nabla f_i(x)$ as a subgradient of f_i at x, and only keep the differentiability condition on the limit function f.

Lemma 4.2. *Assume that* (A_1)-(A_5), (3.5) *and* (3.6) *hold. Then*

$$\hat{M}_n-M_n=\sum_{i=1}^{n}K_n'X_{in}\psi'(\varepsilon_i)\Lambda^{-1}+o_p(1),$$

$$\hat{M}_n^*-M_n^*=\sum_{i=1}^{n}(\Lambda^{-1}\psi(\varepsilon_i))\otimes(K_n'X_{in})+o_p(1). \tag{4.14}$$

Especially for the unrestricted M-*estimate* $\hat{B}_n=S_n^{1/2}\hat{B}$ *of* B_n, *we have*

$$\hat{B}_n-B_n=\sum_{i=1}^{n}X_{in}\psi'(\varepsilon_i)\Lambda^{-1}+o_p(1),$$

$$\hat{B}_n^*-B_n^*=\sum_{i=1}^{n}(\Lambda^{-1}\psi(\varepsilon_i))\otimes X_{in}+o_p(1). \tag{4.15}$$

Proof. Write

$$\bar{M} = M_n + \sum_{i=1}^{n} K_n' X_{in} \psi'(\varepsilon_i) \Lambda^{-1},$$

or

$$\bar{M}^* = M_n^* + \sum_{i=1}^{n} (\Lambda^{-1}\psi(\varepsilon_i)) \otimes (K_n' X_{in}).$$

Since $\bar{M}^* - M_n^*$ has an asymptotic normal distribution, we have

$$\|\bar{M} - M_n\| = \mathrm{O}_p(1). \tag{4.16}$$

By (4.11), it follows that for any $c>0$ and $\delta>0$,

$$\sup_{\|G-\bar{M}\|=\delta} I(\|\bar{M}-M_n\| \leqslant c) \left| \sum_{i=1}^{n} \{\varrho(Y_i - G'K_n'X_{in}) - \varrho(Y_i - \bar{M}'K_n'X_{in})\} - \tfrac{1}{2}(G^* - \bar{M}^*)'(\Lambda \otimes I_{m-q})(G^* - \bar{M}^*) \right| \to 0, \quad \text{in pr.,} \tag{4.17}$$

and that for n large, the event $(\|\bar{M} - M_n\| \leqslant c)$ implies that

$$\inf_{\|G-\bar{M}\|=\delta} \sum_{i=1}^{n} \varrho(Y_i - G'K_n'X_{in}) \geqslant \sum_{i=1}^{n} \varrho(Y_i - \bar{M}'K_n'X_{in}) + \lambda, \tag{4.18}$$

for some $\lambda>0$. By (4.16), (4.18) and the convexity of ϱ, we get

$$P(|\hat{M}_n - \bar{M}| \geqslant \delta) \to 0 \quad \text{as } n \to \infty, \tag{4.19}$$

and (4.14) follows.

Proof of Theorem 3.1. Without loss of generality we assume that $C_0 = 0$. Under H_n we have from (4.6)

$$B_n = K_n M_n + P_n \Theta_n$$

(refer to (4.6), (4.7) and (3.7)). By (4.15) we have

$$\hat{B}_n = B_n + \sum_{i=1}^{n} X_{in} \psi'(\varepsilon_i) \Lambda^{-1} + \mathrm{o}_p(1). \tag{4.20}$$

By (2.4) and (3.1),

$$W_n = \hat{B}'P(P'S_n^{-1}P)^{-1}P'\hat{B} = \hat{B}_n' P_n P_n' \hat{B}_n. \tag{4.21}$$

By (4.6), (4.20) and (4.4) we get

$$P_n'\hat{B}_n = U_n + \Theta_n \tag{4.22}$$

and the theorem follows from (4.21) and (4.22).

Proof of Theorem 3.2. Under H_n, we have $B_n = K_n M_n + P_n \Theta_n$. By (4.14),

$$\|\hat{M}_n - M_n\| = \mathrm{O}_p(1). \tag{4.23}$$

By (4.10) and (4.23),

$$\sum_{i=1}^{n} \{\psi(y_i - \tilde{B}_n' X_{in}) - \psi(\varepsilon_i)\} \otimes X_{in}$$
$$+ (\Lambda \otimes K_n)(\hat{M}_n^* - M_n^*) - (\Lambda \otimes P_n)\Theta_n^* \to 0 \quad \text{in pr.}, \tag{4.24}$$

which implies that

$$\sum_{i=1}^{n} (\psi(Y_i - \tilde{B}_n' X_{in}) - \psi(\varepsilon_i)) \otimes (K_n' X_{in})$$
$$+ (\Lambda \otimes I_{m-q})(\hat{M}_n^* - M_n^*) \to 0 \quad \text{in pr.}, \tag{4.25}$$

and

$$\sum_{i=1}^{n} (\psi(Y_i - \tilde{B}_n' X_{in}) - \psi(\varepsilon_i)) \otimes (P_n' X_{in})$$
$$- (\Lambda \otimes I_q)\Theta_n^* \to 0, \quad \text{in pr.} \tag{4.26}$$

By (4.14) and (4.25),

$$\sum_{i=1}^{n} \psi(Y_i - \tilde{B}_n' X_{in}) \otimes (K_n' X_{in}) \to 0 \quad \text{in pr.}, \tag{4.27}$$

as $n \to \infty$. By (2.2), (2.6), (4.26) and (4.27), noting that $K_n K_n' + P_n P_n' = I_m$, we have

$$\begin{aligned} R_n &= \left(\sum_{i=1}^{n} \psi(Y_i - \tilde{B}_n' X_{in}) X_{in}' \right) \left(\sum_{j=1}^{n} X_{jn} \psi'(Y_j - \tilde{B}_n' X_{jn}) \right) \\ &= \left(\sum_{i=1}^{n} \psi(Y_i - \tilde{B}_n' X_{in}) X_{in}' K_n \right) \left(\sum_{j=1}^{n} K_n' X_{jn} \psi(Y_j - \tilde{B}_n' X_{jn}) \right) \\ &\quad + \left(\sum_{i=1}^{n} \psi(Y_i - \tilde{B}_n' X_{in}) X_{in}' P_n \right) \left(\sum_{j=1}^{n} P_n' X_{jn} \psi(Y_j - \tilde{B}_n' X_{jn}) \right) \\ &= \left(\sum_{i=1}^{n} \psi(\varepsilon_i) X_{in}' P_n + \Lambda \Theta_n' \right) \left(\sum_{j=1}^{n} P_n' X_{jn} \psi'(\varepsilon_j) + \Theta_n \Lambda \right) + o_p(1), \end{aligned} \tag{4.28}$$

and Theorem 3.2 is proved in view of (3.4).

5. Estimation of the nuisance parameters

In practical applications, we need to estimate the nuisance matrix parameters Γ and Λ. A natural estimate of Γ is

$$\hat{\Gamma} = n^{-1} \sum_{i=1}^{n} \psi(Y_i - \hat{B}' X_i) \psi'(Y_i - \hat{B}' X_i), \tag{5.1}$$

where $\hat{B}$ is an M-estimate of B in the model (1.4). To estimate Λ, we take a $p \times p$

nonsingular matrix Z consisting of $\zeta_1,\dots,\zeta_p$ as its columns, take $h=h_n>0$ such that

$$h_n/d_n\to\infty,\quad h_n\to 0\quad\text{and}\quad \liminf_{n\to\infty} nh_n^2>0, \tag{5.2}$$

define

$$\eta_{ik}=\psi(Y_i-\hat{B}'X_i+h\zeta_k)-\psi(Y_i-\hat{B}'X_i-h\zeta_k),$$
$$i=1,\dots,n,\ k=1,\dots,p, \tag{5.3}$$

and use the $p\times p$ matrix

$$\hat{\Lambda}=(2nh)^{-1}\sum_{i=1}^{n}[\eta_{i1},\dots,\eta_{ip}]Z^{-1} \tag{5.4}$$

as an estimate of Λ. We have the following theorem:

Theorem 5.1. *Assume that* (A_1)-(A_5) *are satisfied in the model* (1.4). *Then*

$$\hat{\Gamma}\to\Gamma\quad \textit{in pr. as } n\to\infty \tag{5.5}$$

Furthermore, if (5.2) *also holds, then*

$$\hat{\Lambda}\to\Lambda\quad \textit{in pr. as } n\to\infty. \tag{5.6}$$

Note that Zhao and Chen (1990) gave a proof for the special case of $p=1$. However, the proof for the general case of p is more complicated.

Proof. Put $u=(u_1,\dots,u_p)'$, $v=(v_1,\dots,v_p)'$, $\psi(u)=(\psi_1(u),\dots,\psi_p(u))'$. Write $\Theta=\{\theta=(\theta_1,\dots,\theta_p)': \theta_1,\dots,\theta_p=\pm 1\}$. At first we show that, if $|v_k|\leqslant b/2$ for some $b>0$ and $k=1,\dots,p$, there exists a constant $c>0$ such that

$$c\sum_{\theta\in\Theta}\theta'(\psi(u-b\theta)-\psi(u))\leqslant\psi_1(u+v)-\psi_1(u)$$
$$\leqslant c\sum_{\theta\in\Theta}\theta'(\psi(u+b\theta)-\psi(u)) \tag{5.7}$$

and similar inequalities hold for $\psi_k(u+v)-\psi_k(u)$, $k=2,\dots,p$.

Note that $\theta'(\psi(u+b\theta)-\psi(u))\geqslant 0$ and $\theta'(\psi(u-b\theta)-\psi(u))\leqslant 0$ for any $\theta\in\Theta$.

In fact, by the cyclical monotonicity of ψ (refer to Rockafellar, 1970, p. 238), for any $\theta\in\Theta$ we have

$$v'\psi(u)+(b\theta-v)'\psi(u+v)-b\theta'\psi(u+b\theta)\leqslant 0 \tag{5.8}$$

which implies that

$$(b\theta-v)'(\psi(u+v)-\psi(u))\leqslant b\theta'(\psi(u+b\theta)-\psi(u)) \tag{5.9}$$

and

$$(b\theta+v)'(\psi(u+v)-\psi(u))\geqslant b\theta'(\psi(u-b\theta)-\psi(u)). \tag{5.10}$$

For simplicity we write $\tilde{\theta}=(\theta_1,\dots,\theta_{p-1})'$, $\tilde{u}=(u_1,\dots,u_{p-1})'$, $\tilde{\psi}=(\psi_1,\dots,\psi_{p-1})'$, and sometimes we write $\psi(u')$ for $\psi(u)$. Taking $\theta_p=1$ and -1 in (5.9) we get

$$(b\tilde{\theta}' - \tilde{v}')(\tilde{\psi}(u+v) - \tilde{\psi}(u)) + (b - v_p)(\psi_p(u+v) - \psi_p(u))$$
$$\leqslant b(\tilde{\theta}', 1)\{\psi(\tilde{u}' + b\tilde{\theta}', u_p + b) - \psi(u)\} \tag{5.11}$$

and

$$(b\tilde{\theta}' - \tilde{v}')(\tilde{\psi}(u+v) - \tilde{\psi}(u)) - (b + v_p)(\psi_p(u+v) - \psi_p(u))$$
$$\leqslant b(\tilde{\theta}', -1)\{\psi(\tilde{u}' + b\tilde{\theta}', u_p - b) - \psi(u)\}. \tag{5.12}$$

Multiplying both sides of (5.12) by $(b - v_p)/(b + v_p)$, and adding the inequality so obtained to (5.11), we eliminate $\psi_p(u+v) - \psi_p(u)$ from (5.11) and (5.12), and get

$$(2b/(b+v_p))(b\tilde{\theta}' - \tilde{v}')(\tilde{\psi}(u+v) - \tilde{\psi}(u))$$
$$\leqslant b(\tilde{\theta}', 1)\{\psi(u' + b(\tilde{\theta}', 1)) - \psi(u)\} + b(\tilde{\theta}', -1)\{\psi(u' + b(\tilde{\theta}', -1))$$
$$- \psi(u)\}(b - v_p)/(b + v_p). \tag{5.13}$$

Now it is not difficult to get the second inequality of (5.7) by using the elimination method step by step. The first inequality of (5.7) could be obtained similarly from (5.10).

Without loss of generality, we assume that the true parameter matrix $B = 0$ in the model (1.4). By (4.15) and $B_n = 0$, we have

$$P(\|\hat{B}_n\| \geqslant d_n^{-1/2}) \to 0 \quad \text{as } n \to \infty. \tag{5.14}$$

By (A_4) and the strong law of large numbers,

$$\Gamma_n \triangleq n^{-1} \sum_{i=1}^{n} \psi(\varepsilon_i)\psi'(\varepsilon_i) \to \Gamma = (\gamma_{lm}) \quad \text{a.s.,} \tag{5.15}$$

as $n \to \infty$. Putting $\hat{\Gamma} = (\hat{\gamma}_{lm})$, $\Gamma_n = (\gamma_{lm}^{(n)})$, we have

$$|\hat{\gamma}_{lm} - \gamma_{lm}^{(n)}|^2 = \left| n^{-1} \sum_{i=1}^{n} \{\psi_l(\varepsilon_i - \hat{B}_n' X_{in})\psi_m(\varepsilon_i - \hat{B}_n' X_{in}) - \psi_l(\varepsilon_i)\psi_m(\varepsilon_i)\} \right|^2$$
$$\leqslant n^{-1} \sum_{i=1}^{n} (\psi_l(\varepsilon_i - \hat{B}_n' X_{in}) - \psi_l(\varepsilon_i))^2 \cdot n^{-1} \sum_{i=1}^{n} \psi_m^2(\varepsilon_i - \hat{B}_n' X_{in})$$
$$+ n^{-1} \sum_{i=1}^{n} \psi_l^2(\varepsilon_i) \cdot n^{-1} \sum_{i=1}^{n} (\psi_m(\varepsilon_i - \hat{B}_n' X_{in}) - \psi_m(\varepsilon_i))^2. \tag{5.16}$$

On the event $(\|\hat{B}_n\| < d_n^{-1/2})$, $\|\hat{B}_n' X_{in}\| < d_n^{1/2}$ for each i. By (5.7), there exists a positive constant c such that

$$n^{-1} \sum_{i=1}^{n} (\psi_l(\varepsilon_i - \hat{B}_n' X_{in}) - \psi_l(\varepsilon_i))^2 I(\|\hat{B}_n\| < d_n^{-1/2})$$
$$\leqslant c \max \left\{ n^{-1} \sum_{i=1}^{n} \|\psi(\varepsilon_i + 2d_n^{1/2}\theta) - \psi(\varepsilon_i)\|^2 : \right.$$
$$\left. \theta = (\theta_1, \ldots, \theta_p)', \ \theta_1, \ldots, \theta_p = \pm 1 \right\}. \tag{5.17}$$

By (A_3) and (A_5), for fixed θ we have

$$E\left\{n^{-1}\sum_{i=1}^{n}\|\psi(\varepsilon_i+2d_n^{1/2}\theta)-\psi(\varepsilon_i)\|^2 = E\|\psi(\varepsilon_1+2d_n^{1/2}\theta)-\psi(\varepsilon_1)\|^2\to 0\right. \tag{5.18}$$

as $n\to\infty$. By (5.14) and (5.16)-(5.18), we get

$$\lim_{n\to\infty}|\hat{\gamma}_{lm}-\gamma_{lm}^{(n)}|=0 \quad \text{in pr., for } l,m=1,\ldots,p, \tag{5.19}$$

which implies (5.5) in view of (5.15).

Now we proceed to prove (5.6). To this end, we prove that for any $c>0$,

$$(nh)^{-1}\sum_{i=1}^{n}(\psi_l(\varepsilon_i-\hat{B}_n'X_{in}+h\zeta_k)-\psi_l(\varepsilon_i+h\zeta_k))I(\|\hat{B}_n\|\leqslant c)\to 0$$
$$\text{in pr., } l=1,\ldots,p. \tag{5.20}$$

By (5.7), in order to prove (5.20), it is enough to prove that for each fixed $\theta\in\Theta$,

$$T_n\triangleq(nh)^{-1}\sum_{i=1}^{n}\theta'(\psi(\varepsilon_i+2cd_n\theta+h\zeta_k)-\psi(\varepsilon_i+h\zeta_k))\to 0$$
$$\text{in pr., as } n\to\infty. \tag{5.21}$$

By (A_3), (A_5) and (5.2), we have

$$\begin{aligned}\operatorname{Var} T_n&\leqslant(nh^2)^{-1}E[\theta'(\psi(\varepsilon_1+2cd_n\theta+h\zeta_k)-\psi(\varepsilon_1+h\zeta_k))]^2\\&\leqslant(nh^2)^{-1}\|\theta\|^2E\|\psi(\varepsilon_1+2cd_n\theta+h\zeta_k)-\psi(\varepsilon_1+h\zeta_k)\|^2\to 0.\end{aligned} \tag{5.22}$$

On the other hand, by (A_2), (A_5) and (5.2), we get

$$ET_n=h^{-1}\theta'E(\psi(\varepsilon_1+2cd_n\theta+h\zeta_k)-\psi(\varepsilon_1+h\zeta_k))\to 0. \tag{5.23}$$

By (5.22) and (5.23), we get (5.21) and (5.20). Noting that $\|\hat{B}_n\|=O_p(1)$, we have

$$(nh)^{-1}\sum_{i=1}^{n}(\psi(\varepsilon_i-\hat{B}_n'X_{in}+h\zeta_k)-\psi(\varepsilon_i+h\zeta_k))\to 0 \quad \text{in pr.} \tag{5.24}$$

In the same way,

$$(nh)^{-1}\sum_{i=1}^{n}(\psi(\varepsilon_i-\hat{B}_n'x_{in}-h\zeta_k)-\psi(\varepsilon_i-h\zeta_k))\to 0 \quad \text{in pr.} \tag{5.25}$$

By (A_3) and (5.2), for $m=1,\ldots,p$.

$$\begin{aligned}&\operatorname{Var}\left\{(nh)^{-1}\sum_{i=1}^{n}(\psi_m(\varepsilon_i\pm h\zeta_k)-\psi_m(\varepsilon_i))\right\}\\&\quad\leqslant(nh^2)^{-1}E(\psi_m(\varepsilon_1\pm h\zeta_k)-\psi_m(\varepsilon_1))^2\to 0 \quad \text{as } n\to\infty.\end{aligned} \tag{5.26}$$

On the other hand, by (A_2) and $h_n \to 0$, we have

$$(2nh)^{-1}\sum_{i=1}^{n} E(\psi(\varepsilon_i + h\zeta_k) - \psi(\varepsilon_i - h\zeta_k))$$
$$= (2h)^{-1}E(\psi(\varepsilon_1 + \zeta_k) - \psi(\varepsilon_1 - h\zeta_k)) \to \Lambda\zeta_k, \quad k = 1, \dots, p. \tag{5.27}$$

By (5.26) and (5.27), we have

$$(2nh)^{-1}\sum_{i=1}^{n} (\psi(\varepsilon_i + h\zeta_k) - \psi(\varepsilon_i - h\zeta_k)) \to \Lambda\zeta_k \quad \text{in pr.}, \tag{5.28}$$

for $k = 1, \dots, p$. By (5.24), (5.25), (5.28) and (5.3), it follows that

$$\hat{\Lambda}Z = (2nh)^{-1}\sum_{i=1}^{n} [\eta_{i1}, \dots, \eta_{ip}] \to \Lambda Z \quad \text{in pr.}, \tag{5.29}$$

and (5.6) is obtained. Now Theorem 5.1 is proved.

Note 1. In estimating Λ and proving the consistency of the estimate, we have not made any additional assumptions on ϱ. The only property used is its convexity. If, however, ϱ is twice differentiable, other estimates are possible, as in the case of the least distances estimate considered by Bai, Chen, Miao and Rao (1990).

Note 2. A referee remarks that Theorem 5.1 can be proved by applying the convexity lemma in a recent paper by Pollard (1991). It is true, but the detailed proof given by us using similar ideas will be of help in solving similar problems. Pollard's paper which contains results similar to the earlier papers by Bai, Rao and Yin (1990) and Chen, Bai, Zhao and Wu (1990) was not available to us when our paper was submitted for publication.

Note added in proof. This work can be extended to a more general case where $\varrho = \varrho^{(1)} - \varrho^{(2)}$ is a difference of two p-variate convex functions $\varrho^{(1)}$ and $\varrho^{(2)}$ with $\psi(u) = \psi^{(1)}(u) - \psi^{(2)}(u)$ being the difference of their subgradients at u. Assume that (A_1)-(A_5) are satisfied with ϱ and ψ in (A_1)-(A_5) being replaced by $\varrho^{(1)}, \psi^{(1)}$ and $\varrho^{(2)}, \psi^{(2)}$. We can construct the same test statistics with $\varrho = \varrho^{(1)} - \varrho^{(2)}$ and $\psi = \psi^{(1)} - \psi^{(2)}$ as before. It can be shown that, if the above conditions are met, Theorems 3.1, 3.2 and 5.1 are still valid. In this context, the minimizer of the relevant function could be taken as its some local minimizer having some properties. For the details, refer to Bai, Rao and Wu (1992).

References

Bai, Z.D., C.R. Rao and Y. Wu (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica*, **2**, 237-254.

Bai, Z.D., C.R. Rao and Y. Wu (1992). A note on M-estimation of multivariate linear regression. Technical Report No. 92-14, Center for Multivariate Analysis, Penn State University. To appear in *Statistica Sinica*.

Bai, Z.D., X.R. Chen, B.Q. Miao and C.R. Rao (1990). Asymptotic theory of least distances estimates in multivariate analysis. *Statistics*, **21**, 503–519.

Bai, Z.D., C.R. Rao and Y.Q. Yin (1990). Least absolute deviations analysis of variance. *Sankhyā*, **52**, 166–177.

Chen, X.R., Z.D. Bai, L.C. Zhao and Y. Wu (1990). Asymptotic normality of minimum L_1-norm estimates in linear models. *Science in China Series A*, **33**, 1311–1328.

Heiler, S. and R. Willers (1988). Asymptotic normality of R-estimates in the linear model. *Statistics*, **19**, 173–184.

Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Ann. Instit. Statist. Math.*, **25**, 1–26.

Jurečková, J. (1983). Robust estimators of location and regression parameters and their second order asymptotic relations. *Trans. 9th. Prague Confer. Inform. Theor. Stat. Dec. Func. Ran. Proc.*, pp. 19–32.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, **7**, 186–199.

Rao, C.R. (1948). Tests of significance in multivariate analysis. *Biometrika*, **35**, 58–79.

Rao, C.R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

Rao, C.R. (1973). Linear Statistical Inference and Its Applications, Second Edition. Wiley, New York.

Rockafellar, R.T. (1970). *Convex Analysis.* Princeton University Press, Princeton, NY.

Schrader, R.M. and T.P. Hettmansperger (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika*, **67**, 93–101.

Sen, P.K. (1982). On M tests in linear models. *Biometrika*, **69**, 245–248.

Singer, J.M. and P.K. Sen (1985). M-methods in multivariate linear models. *J. Multivariate Anal.*, **17**, 168–184.

Zhao, L.C. and X.R. Chen (1990). Asymptotic behavior of M-test statistics in linear models. Technical Report no. 90-37, Center for Multivariate Analysis, Penn State University.