主論文

Some estimators of covariance matrix in multivariate nonparametric regression and their applications

（多変量ノンパラメトリック回帰分析における共分散行列の推定量およびその応用）

Megu OHTAKI

To be published in
Hiroshima Mathematical Journal, Vol. 20, No. 1, 1990

SOME ESTIMATORS OF COVARIANCE MATRIX IN MULTIVARIATE NONPARAMETRIC REGRESSION AND THEIR APPLICATIONS

Megu Ohtaki

(Recieved January 20, 1989)

CONTENTS

1. Introduction
2. A class of estimators
3. Upper bounds for biases
4. Efficiency
5. Asymptotic properties
6. Some special cases when q = 2
7. Testing goodness of fit of linear models
8. Robust estimators of diagonal elements of Σ
9. Appendix. Covariances of some quadratic forms

1. Introduction

Consider the regression problem on a set of p response variables $\mathbf{y} = (y_1, \cdots, y_p)'$ and a set of q explanatory variables $\mathbf{x} = (x_1, \cdots, x_q)'$. Let $(\mathbf{y}_i = (y_{i1}, \cdots, y_{ip})'; \mathbf{x}_i = (x_{i1}, \cdots x_{iq})')$, $i = 1, \cdots, n$, be the n observations on $(\mathbf{y}; \mathbf{x})$. The regression model assumed is

$$\mathbf{y}_i = \boldsymbol{\eta}(\mathbf{x}_i) + \boldsymbol{\varepsilon}_i, \tag{1.1}$$

where $\boldsymbol{\eta} = (\eta_1, \cdots \eta_p)': R^q \to R^p$ is a function of $\mathbf{x}$ whose shape is unknown but its smoothness is presumed, and the errors $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ip})'$, $i = 1, \cdots, n$, are independently and identically distributed with mean $\mathbf{0}$ and unknown covariance matrix $\Sigma = [\sigma_{j\ell}]_{p\times p}$. Writing this model in matrix form, we have

$$Y = \P + \mathcal{E}, \tag{1.2}$$

where

$$Y = \begin{bmatrix} \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_n' \end{bmatrix} = [\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(p)}],$$

$$\P = \begin{bmatrix} \boldsymbol{\eta}_1' \\ \vdots \\ \boldsymbol{\eta}_n' \end{bmatrix} = [\boldsymbol{\eta}^{(1)}, \cdots, \boldsymbol{\eta}^{(p)}],$$

and $$\mathcal{E} = \begin{bmatrix} \underset{\sim}{\varepsilon}_1' \\ \vdots \\ \underset{\sim}{\varepsilon}_n' \end{bmatrix} = [\underset{\sim}{\varepsilon}^{(1)}, \cdots, \underset{\sim}{\varepsilon}^{(p)}].$$

The measurements $\underset{\sim}{x}_i$, $i = 1, \cdots, n$, which are called the design points are expressed as

$$X = \begin{bmatrix} \underset{\sim}{x}_1' \\ \vdots \\ \underset{\sim}{x}_n' \end{bmatrix} = [\underset{\sim}{x}^{(1)}, \cdots, \underset{\sim}{x}^{(q)}].$$

It is assumed that $x_{i1} = 1$, for $i = 1, \cdots, n$, i.e., $\underset{\sim}{x}^{(1)} = \underset{\sim}{1}_n$ and $\mathrm{rank}(X) = q \leq n$.

The regression analysis usually involves two important problems; making inferences about the regression surface $\underset{\sim}{\eta}$ and estimating the covariance matrix Σ. These problems are closely related. It is easily seen that a good estimator of $\underset{\sim}{\eta}$ immediately yields a good one of Σ. Conversely, once an adequate estimator of Σ is available, it will provide helpful information to explore a good estimator of $\underset{\sim}{\eta}$. When a vaild parametric model for $\underset{\sim}{\eta}$ is at hand, some least squares technique will yield a good result. However, in practical situation of data analysis it is often difficult to chose a valid parametric model especially when q or p is large (see, e.g., Cleveland and Devlin[4], Silverman[20], Rice[16], Ohtaki[13]). For such a situation, it may be a good strategy to start the analysis by estimating Σ rather than $\underset{\sim}{\eta}$ nonparametrically.

The simplest nonparametric estimator of Σ may be constructed by making use of replicated observations. Suppose that there are g distinct sets of replicated observations $\{(\underset{\sim}{y}_{\ell t}, \underset{\sim}{x}_{\ell}) \mid 1 \le t \le m_{\ell}\}$, ℓ = 1, $\cdots$, g, in data. Then, an unbiased estimator of Σ is given by

$$(1.3) \qquad \hat{\Sigma}_{PE} = \{\sum_{\ell=1}^{g}(m_{\ell}-1)\}^{-1}\sum_{\ell=1}^{g}\sum_{t=1}^{m_{\ell}}(\underset{\sim}{y}_{\ell t}-\bar{\underset{\sim}{y}}_{\ell \cdot})(\underset{\sim}{y}_{\ell t}-\bar{\underset{\sim}{y}}_{\ell \cdot})',$$

where $\bar{\underset{\sim}{y}}_{\ell \cdot} = m_{\ell}^{-1}\sum_{t=1}^{m_{\ell}} \underset{\sim}{y}_{\ell t}$. This estimator $\hat{\Sigma}_{PE}$ is refered to as (Multivariate) ***Pure error mean square*** (PEMS) estimator (see, e.g., Draper and Smith[7, Section 1.5] Weisberg[23, Section 4.3]). Unfortunately, this estimator often lose its effectiveness because no or very few replicated observations are available in most data. Daniel and Wood [5] suggested the use of an approximate PEMS estimator. Their idea is to use a clustering algorithm to find the cases that are *almost replicates*, and use the variation of the responses for the *almost replicates*. An interesting application of their idea to logistic regression was given by Landwehr et al.[12]. Recently, Gasser et al.[8] and Ohtaki[14] have proposed a class of estimator of variance in univariate one-dimensional nonparametric regression model, i.e., the case of p = 1 and q = 2. Some properties of the estimators have been studied by Gasser et al.[8], Ohtaki[14] and Buckley et al.[3]. In this paper these results in univariate cases are extended to ones in multivariate situations. The outline of this paper is as follows: In Section 2 we introduce a class of nonparametric estimators. The biases of those estimators are studied and their upper bounds are given in Section 3. In Section 4 we derive the exact formulas of covariance matrices of the estimators, and assess the efficiency by comparing with the best linear unbiased estimator under the linear regression model. In Section 5 we investigate some asymptotic behaviors of the estimators and show the sufficient conditions for consistency or asymptotic normality. In Section 6 we consider the case when $q = 2$ in detail; we provide a multivariate extention of the estimators which were proposed in univariate regression model by Gasser et al.[8] and by Ohtaki[14], and show that the newly obtained estimators become a natural extention of PEMS estimator. In Section 7 we propose a new type of test statistics for assessing goodness of fit of linear models, and prove that the asymptotic null distribution of the criterion is N(0,1) under some mild regularity conditions. In Section 8, using the idea due to Rousseeuw[17], we construct a robust alternative to the diagonal elements of covariance matrix, and show that the robust estimator will have a positive breakdown point in some situation.

2. A class of estimators

Suppose that there is a subset K of $\{1, \cdots, n\}$ such that every member i of K has an index-set N_i which specifies a neighborhood of the design point $\underset{\sim}{x}_i$, $\{\underset{\sim}{x}_j \mid j \in N_i\}$. Here it is assumed that $i \notin N_i$. This means that our estimation is based on the cross-varidation technique which will make the resulting estimate of covariance matrix more stable. Let $\hat{\underset{\sim}{y}}_i$ $(i \in K)$ be a linear predictor of $\underset{\sim}{y}_i = (y_{i1}, \cdots, y_{ip})'$ which is based only on the neighborhood $\{(\underset{\sim}{y}_j;\ \underset{\sim}{x}_j) \mid j \in N_i\}$. We write such a predictor as

$$\hat{\underset{\sim}{y}}_i = Y' \underset{\sim}{w}_i \tag{2.1}$$

where $\underset{\sim}{w}_i$ is an n-component vector whose jth component w_{ij} is nonzero only when $j \in N_i$. As for the errors $\hat{\underset{\sim}{r}}_i = \hat{\underset{\sim}{y}}_i - \underset{\sim}{y}_i$, it is easily seen that

$$E[\underset{\sim}{r}_i \underset{\sim}{r}_i'] = c_i^{-2}\Sigma + \underset{\sim}{\xi}_i \underset{\sim}{\xi}_i', \qquad i \in K, \tag{2.2}$$

where $\underset{\sim}{\xi}_i = E[\underset{\sim}{r}_i] = \P'(\underset{\sim}{w}_i - \underset{\sim}{\delta}_i)$, $\underset{\sim}{\delta}_i = (\delta_{i1}, \cdots, \delta_{in})'$, δ_{ij} is the Kronecher delta, $c_i^2 = 1/(1 + \|\underset{\sim}{w}_i\|^2)$ and $\|\underset{\sim}{w}_i\| = \sqrt{\underset{\sim}{w}_i'\underset{\sim}{w}_i}$.

The result (2.2) suggests that an estimator of Σ may be obtained obtained through averaging $c_i^2 \underset{\sim}{r}_i \underset{\sim}{r}_i'$, $i \in K$. Adopting the set of weights $\{c_i^2\}$, we propose the following class of estimators of Σ:

$$\hat{\Sigma}_{\mathcal{N}} = \left(\sum_{i \in K} c_i^2\right)^{-1} \sum_{i \in K} c_i^4 \underset{\sim}{r}_i \underset{\sim}{r}_i'. \tag{2.3}$$

The (j, ℓ)-element of $\hat{\Sigma}_{\mathcal{N}}$ is expressed as

$$\hat{\sigma}_{\mathcal{N}}(j,\ell) = \left(\sum_{i \in K} c_i^2\right)^{-1} \sum_{i \in K} c_i^4 r_{ij} r_{i\ell}, \qquad 1 \le j,\ \ell \le p. \tag{2.4}$$

REMARK 2.1. The estimator $\hat{\Sigma}_{\mathcal{N}}$ of (2.3) is expressed in a matrix notation as

$$\hat{\Sigma}_{\mathcal{N}} = (\mathrm{tr} V_{\mathcal{N}})^{-1} Y' V_{\mathcal{N}} Y, \tag{2.5}$$

where

(2.6) $$V_{\mathcal{N}} = \sum_{i \in K} c_i^4 (\underset{\sim}{w}_i - \underset{\sim}{\delta}_i)(\underset{\sim}{w}_i - \underset{\sim}{\delta}_i)'.$$

The matrix $V_{\mathcal{N}}$ is non-negative definite and its (α,β)-element $v_{\alpha\beta}$ is expressed as

(2.7) $$v_{\alpha\beta} = c_\alpha^4 \delta_{\alpha\beta} I_{\{\alpha \in K\}} - c_\alpha^4 w_{\alpha\beta} I_{\{\alpha \in K\}} - c_\beta^4 w_{\beta\alpha} I_{\{\beta \in K\}} + \sum_{\gamma \in K} c_\gamma^4 w_{\gamma\alpha} w_{\gamma\beta},$$

where $I_{\{E\}} = 1$ if the statement E is true, and 0 otherwise.

It is possible to use another sets of weights instead of $\{c_i^2\}$ in averaging $c_i^2 \underset{\sim}{r}_i \underset{\sim}{r}_i'$ $(i \in K)$. For example, homogeneous weights n_K^{-1} (n_K is the total number of elements of K) was adopted by Gasser et al.[8]. An advantage for using $\{c_i^2\}$ as a set of weights is that the resulting estimator of Σ becomes a natural extension of PEMS estimator. This will be shown in EXAMPLE 2.1.

Two important special cases of the estimator (2.3) are given in the following examples.

EXAMPLE 2.1(*locally uniform weight* (LUW) estimator). Let the weight-vector $\underset{\sim}{w}_i$ in (2.1) be an n-component vector having the jth element

(2.8) $$w_{ij} = \begin{cases} 1/n_i, & \text{if } j \in N_i, \\ 0, & \text{if } j \notin N_i, \end{cases}$$

where n_i denotes the number of elements in N_i. Then, $\hat{\underset{\sim}{y}}_i = \bar{\underset{\sim}{y}}_{(i)} = \sum_{j \in N_i} \underset{\sim}{y}_j / n_i$, so that the resulting estimator can be expressed as

$$\hat{\Sigma}_{\mathfrak{U}} = \left(\sum_{i\in K} \frac{n_i}{n_i+1}\right)^{-1} \sum_{i\in K} \left(\frac{n_i}{n_i+1}\right)^2 (\hat{y}_i - y_i)(\hat{y}_i - y_i)'$$

$$= \left(\sum_{i\in K} \frac{n_i}{n_i+1}\right)^{-1} \sum_{i\in K} (\bar{y}_{i\cdot} - y_i)(\bar{y}_{i\cdot} - y_i)',$$

where $\bar{y}_{i\cdot} = \left(y_i + \sum_{j\in N_i} y_j\right)/(n_i+1)$, $i \in K$. This estimator will be refered to as a *locally uniform weight* (LUW) estimator.

Consider the situation where every ith set $\{x_j \mid j \in N_i \text{ or } j = i\}$ $(i \in K)$ consists of m_i replicates and there are g distinct design points. Then, using the notation in (1.3), we have

$$\sum_{i\in K} \frac{n_i}{n_i+1} = \sum_{\ell=1}^{g} \sum_{t=1}^{m_\ell} (m_\ell - 1)/m_\ell = \sum_{\ell=1}^{g} (m_\ell - 1),$$

and

$$\sum_{i\in K} (y_i - \bar{y}_{i\cdot})(y_i - \bar{y}_{i\cdot})' = \sum_{\ell=1}^{g} \sum_{t=1}^{m_\ell} (y_{\ell t} - \bar{y}_{\ell\cdot})(y_{\ell t} - \bar{y}_{\ell\cdot})'.$$

This implies that $\hat{\Sigma}_{\mathfrak{U}} = \hat{\Sigma}_{PE}$. Thus, we see that the PEMS estimator defined by (1.3) is a special case of LUW estimator. Even though PEMS estimator is generally biased unless underlying regression function is exactly constant, it has a computational convenience and may also provide satisfactory information on Σ in some practical regression situations.

EXAMPLE 2.2(*locally linear weight* (LLW) estimator). It may be noted that the locally linear model may reduce effectively the possible bias in the resulting estimator of Σ, as Stone[22] has suggested in general context of nonparametric regression. Let $\hat{\underset{\sim}{y}}_i = \hat{B}_i' \underset{\sim}{x}_i$, where $\hat{B}_i = [b_{j\ell}^{(i)}]$ is the q × p matrix which minimizes

$$\mathrm{tr}[(Y - XB_i)' D_i (Y - XB_i)]$$

$$= \sum_{j \in N_i} (\underset{\sim}{y}_j - B_j' \underset{\sim}{x}_i)'(\underset{\sim}{y}_j - B_j' \underset{\sim}{x}_i),$$

where $D_i = \mathrm{diag}[d_1^{(i)}, \cdots, d_n^{(i)}]$ and

$$d_j^{(i)} = \begin{cases} 1, & \text{if } j \in N_i, \\ 0, & \text{if } j \notin N_i. \end{cases}$$

This linear predictor is based on the least squares estimators in fitting a linear regression model to the data $\{(\underset{\sim}{y}_j;\ \underset{\sim}{x}_j) \mid j \in N_i\}$. Then the predictor is written in the form $\hat{\underset{\sim}{y}}_i = Y' \underset{\sim}{w}_i$, and its weight-vector is given by

$$\underset{\sim}{w}_i = D_i X (X' D_i X)^- \underset{\sim}{x}_i, \quad i \in K, \tag{2.9}$$

where A^- denotes a general inverse of A. We note that $\underset{\sim}{w}_i' \underset{\sim}{1}_n = 1$, since $\underset{\sim}{x}^{(1)} = \underset{\sim}{1}_n$. The resulting estimator of Σ will be refered to as a *locally linear weight* (LLW) estimator and denoted by $\hat{\Sigma}_{\mathcal{L}}$. Using a few algebra, we obtain that $\hat{\Sigma}_{\mathcal{L}} = \hat{\Sigma}_{PE}$ when every ith set $\{\underset{\sim}{x}_j \mid j \in N_i$ or $j = i\}$ consists of only replicated design points which are identical to $\underset{\sim}{x}_i$ $(i \in K)$. Thus, we see that the LUW estimator is also a natuaral extension of PEMS estimator.

3. Upper bounds for biases

Let $\hat{\Sigma}_{\mathcal{N}}$ be a nonparametric estimator of Σ definded by (2.3). A few calculation yields the following formula for the expectation of $\hat{\Sigma}_{\mathcal{N}}$:

$$(3.1) \qquad E[\hat{\Sigma}_{\mathcal{N}}] = \Sigma + \left(\sum_{i \in K} c_i^2\right)^{-1} \sum_{i \in K} c_i^4 \xi_i \xi_i'.$$

where

$$(3.2) \qquad \xi_i = E[r_i] = E[\hat{y}_i - y_i] = \P' w_i.$$

It is easy to see that the second term of (3.1) is a non-negative definite matrix, and hence the estimator $\hat{\Sigma}_{\mathcal{N}}$ of Σ is always positively biased unless $\xi_i = 0$ for all $i \in K$. The following LEMMAS 3.1 and 3.2 are foundamental in obtaining upper bounds for biases of two estimators $\hat{\Sigma}_{\mathcal{U}}$ and $\hat{\Sigma}_{\mathcal{L}}$.

LEMMA 3.1. Suppose that a function f: $R^q \to R^1$ is differentiable. Let $\Delta_i = w_i' f - f(x_i)$, where $f = (f(x_1), \cdots, f(x_n))'$ and the weight-vector w_i is given by (2.8). Then

$$(3.3) \qquad |\Delta_i| \leq \psi_f d_i,$$

where $\psi_f = \sup_{t} \left\{ \sum_{\ell=1}^{q} \left(\frac{\partial}{\partial x_\ell} f(x) \big|_{x=t} \right)^2 \right\}^{\frac{1}{2}}$ and $d_i = \max_{j \in N_i} \| x_j - x_i \|$.

Proof. Using a Taylor expansion of f about x_i, we have

$$|f(\underset{\sim}{x}) - f(\underset{\sim}{x}_i)| \le \psi_f \|\underset{\sim}{x} - \underset{\sim}{x}_i\|.$$

Since $\Delta_i = (1/n_i) \sum_{j \in N_i} \{f(\underset{\sim}{x}_j) - f(\underset{\sim}{x}_i)\}$, we have

$$|\Delta_i| \le \psi_f n_i^{-1} \sum_{j \in N_i} \|\underset{\sim}{x}_j - \underset{\sim}{x}_i\| \le \psi_f d_i.$$

LEMMA 3.2. Suppose that a function $f : R^q \to R^1$ is twice differentiable. Let $\Delta_i = \underset{\sim}{w}_i' \underset{\sim}{f} - f(\underset{\sim}{x}_i)$, where $\underset{\sim}{f} = (f(\underset{\sim}{x}_1), \cdots, f(\underset{\sim}{x}_n))'$ and the weight-vector $\underset{\sim}{w}_i$ is given by (2.9). Then

$$(3.4) \qquad |\Delta_i| \le \frac{1}{2} \gamma_f \sqrt{n_i} \|\underset{\sim}{w}_i\| d_i^2,$$

where $\gamma_f = \sup_{\underset{\sim}{z}} \sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} |\underset{\sim}{u}' H_{\underset{\sim}{z}} \underset{\sim}{u}|$, $H_{\underset{\sim}{z}}$ is the Hessian of f at $\underset{\sim}{x} = \underset{\sim}{z}$, $\|\underset{\sim}{w}_i\| = \sqrt{\underset{\sim}{w}_i' \underset{\sim}{w}_i}$ and n_i is the number of elements in N_i.

Proof. Using a Taylor expansion of f about $\underset{\sim}{x}_i$, we have

$$f(\underset{\sim}{x}) = f(\underset{\sim}{x}_i) + (\underset{\sim}{x} - \underset{\sim}{x}_i)' \underset{\sim}{b}_i + \frac{1}{2} (\underset{\sim}{x} - \underset{\sim}{x}_i)' H_i (\underset{\sim}{x} - \underset{\sim}{x}_i),$$

where $\underset{\sim}{b}_i' = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_q} \right) \Big|_{\underset{\sim}{x} = \underset{\sim}{x}_i}$, $H_i = H_{\underset{\sim}{z}_i}$ and $\underset{\sim}{z}_i = \tau_i \underset{\sim}{x} + (1 - \tau_i) \underset{\sim}{x}_i$ for some τ_i in (0, 1), i = 1, ⋯, n. Let $\underset{\sim}{R}_i' = \frac{1}{2} \Big((\underset{\sim}{x}_1 - \underset{\sim}{x}_i)' H_1 (\underset{\sim}{x}_1 - \underset{\sim}{x}_i), \cdots, (\underset{\sim}{x}_n - \underset{\sim}{x}_i)' H_n (\underset{\sim}{x}_n - \underset{\sim}{x}_i) \Big)$. Since $\underset{\sim}{f} = f(\underset{\sim}{x}_i) \underset{\sim}{1}_n + (X - \underset{\sim}{1}_n \underset{\sim}{x}_i') \underset{\sim}{b}_i + \underset{\sim}{R}_i$ and $\underset{\sim}{w}_i' \underset{\sim}{1}_n = 1$, we have

$$\underset{\sim}{w}_i' \underset{\sim}{f} = f(\underset{\sim}{x}_i) + \underset{\sim}{x}_i' (X' D_i X)^{-} X' D_i \underset{\sim}{R}_i.$$

Hence, $\Delta_i = \underset{\sim}{w}_i'\underset{\sim}{f} - f(\underset{\sim}{x}_i) = \underset{\sim}{x}_i'(X'D_iX)^-X'D_i\underset{\sim}{R}_i$, $i \in K$. Note that the largest eigenvalue of $D_i\underset{\sim}{R}_i\underset{\sim}{R}_i'D_i$ can be evaluated as follows:

$$\sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} \underset{\sim}{u}'D_i\underset{\sim}{R}_i\underset{\sim}{R}_i'D_i\underset{\sim}{u} = \frac{1}{4}\sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} \{\sum_{j\in N_i} u_j(\underset{\sim}{x}_j - \underset{\sim}{x}_i)'H_j(\underset{\sim}{x}_j - \underset{\sim}{x}_i)\}^2$$

$$\leq \frac{1}{4}\,\mathrm{tr}D_i \sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} \sum_{j\in N_i} u_j^2\{(\underset{\sim}{x}_j - \underset{\sim}{x}_i)'H_j(\underset{\sim}{x}_j - \underset{\sim}{x}_i)\}^2$$

$$\leq \frac{1}{4}\,n_i \sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} \Big[\max_{1\leq k\leq n}\{\sup_{\underset{\sim}{v}'\underset{\sim}{v}=1}(\underset{\sim}{v}'H_k\underset{\sim}{v})^2$$

$$\times \sum_{j\in N_i} u_j^2(\underset{\sim}{x}_j - \underset{\sim}{x}_i)'(\underset{\sim}{x}_j - \underset{\sim}{x}_i)\}\Big]^2$$

$$\leq \frac{1}{4}\,n_i \sup_{\underset{\sim}{x}}\sup_{\underset{\sim}{u}'\underset{\sim}{u}=1}(\underset{\sim}{u}'H_{\underset{\sim}{x}}\underset{\sim}{u})^2 d_i^4$$

$$\leq \frac{1}{4}\,\gamma_f^2\, n_i d_i^4.$$

Hence, we obtain

$$\Delta_i^2 = \underset{\sim}{x}_i'(X'D_iX)^-X'D_i(D_i\underset{\sim}{R}_i\underset{\sim}{R}_i'D_i)D_iX(X'D_iX)^-\underset{\sim}{x}_i$$

$$\leq \frac{1}{4}\,\gamma_f^2\, n_i d_i^4\, \underset{\sim}{x}_i'(X'D_iX)^-\underset{\sim}{x}_i,$$

$$= \frac{1}{4}\,\gamma_f^2 n_i \|\underset{\sim}{w}_i\|^2 d_i^4.$$

Applying LEMMAS 3.1 and 3.2 for (3.1), we obtain the following theorems:

THEOREM 3.1. Let $\hat{\Sigma}_{\mathcal{U}} = [\hat{\sigma}_{\mathcal{U}}(j,\ell)]$ be a LUW estimator of Σ. Supppose that the jth and the ℓth components η_j and η_ℓ of the regression function $\underset{\sim}{\eta}$ are differentiable, and that

$$\psi_\alpha = \sup_{\underset{\sim}{t}} \left\{ \sum_{k=1}^{q} \left(\frac{\partial}{\partial x_k} \eta_\alpha(\underset{\sim}{x}) |_{\underset{\sim}{x}=\underset{\sim}{t}} \right)^2 \right\}^{\frac{1}{2}} < +\infty, \qquad \alpha = j,\ \ell.$$

Then

$$|E[\hat{\sigma}_{\mathcal{U}}(j,\ell)] - \sigma_{j\ell}| \leq \psi_j \psi_\ell h_{\mathcal{U}}, \tag{3.5}$$

where

$$h_{\mathcal{U}} = \left(\sum_{i \in K} \frac{n_i}{n_i + 1} \right)^{-1} \sum_{i \in K} \left(\frac{n_i}{n_i + 1} d_i \right)^2. \tag{3.6}$$

COROLLARY 3.1. The (j,ℓ)-element $\hat{\sigma}_{\mathcal{U}}(j,\ell)$ of $\hat{\Sigma}_{\mathcal{U}}$ is unbiased if the jth or the ℓth component of the regression function $\underset{\sim}{\eta}$ is exactly constant; therefore, $\hat{\Sigma}_{\mathcal{U}}$ is unbiased if $\underset{\sim}{\eta}$ is a constant function with respect to $\underset{\sim}{x}$.

THEOREM 3.2. Let $\hat{\Sigma}_{\mathcal{L}} = [\hat{\sigma}_{\mathcal{L}}(j,\ell)]$ be a LLW estimator of Σ. Supppose that the jth and the ℓth components η_j and η_ℓ of the regression function $\underset{\sim}{\eta}$ are twice differentiable, and that

$$\gamma_\alpha = \sup_{\underset{\sim}{x}} \sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} |\underset{\sim}{u}' H_{\underset{\sim}{x}}^{(\alpha)} \underset{\sim}{u}| < +\infty, \qquad \alpha = j,\ \ell,$$

where $H_{\underset{\sim}{x}}^{(\alpha)}$ is the Hessian of η_α for $\alpha = j,\ \ell$. Then

$$|E[\hat{\sigma}_{\mathcal{L}}(j,\ell)] - \sigma_{j\ell}| \le \gamma_j\gamma_\ell h_{\mathcal{L}}, \tag{3.7}$$

where

$$h_{\mathcal{L}} = \frac{1}{4}\left(\sum_{i\in K} c_i^2\right)^{-1} \sum_{i\in K} c_i^4 n_i \|\underset{\sim}{w}_i\|^2 d_i^4. \tag{3.8}$$

COROLLARY 3.2. The (j,ℓ)-element $\hat{\sigma}_{\mathcal{L}}(j,\ell)$ of $\hat{\Sigma}_{\mathcal{L}}$ is unbiased if the jth or the ℓth component of the regression function $\underset{\sim}{\eta}$ is exactly linear; therefore, $\hat{\Sigma}_{\mathcal{L}}$ is unbiased if $\underset{\sim}{\eta}$ is a linear function with respect to $\underset{\sim}{x}$.

4. Efficiency

In this section, we assume that the distribution of $\underset{\sim}{\varepsilon}_i$ $(i = 1, \cdots, n)$ have finite fourth moments about $\underset{\sim}{0}$. To give an unified expression for all third or fourth moments, we use the following notation:

$$\mu_3(j,k,\ell) = E[\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{i\ell}], \tag{4.1}$$

$$\mu_4(j,k,\ell,m) = E[\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{i\ell}\varepsilon_{im}], \tag{4.2}$$

for $i = 1, \cdots, n$, for $1 \le j, k, \ell, m \le p$. First we give a general expression for the covariances of linear functions of $\hat{\Sigma}_{\mathcal{N}}$.

THEOREM 4.1. Let $\hat{\Sigma}_{\mathcal{N}}$ be the estimator of Σ defined by (2.3). Suppose that $\underset{\sim}{\varepsilon}_1, \cdots, \underset{\sim}{\varepsilon}_n$ are independently distributed with finite third and fourth moments given by (4.1) and (4.2). If $A = [a_{jk}]$ and $B = [b_{jk}]$ are any $p \times p$ symmetric matrices, then

(4.3) $$\mathrm{Cov}[\mathrm{tr}(A\hat{\Sigma}_N),\ \mathrm{tr}(B\hat{\Sigma}_N)]$$

$$= (\mathrm{tr}V_N)^{-2}[\underset{\sim}{v}_N'\underset{\sim}{v}_N\{\sum_j \sum_k \sum_\ell \sum_m a_{jk}b_{\ell m}\mu_4(j,k,\ell,m)$$

$$- \mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma) - 2\mathrm{tr}(A\Sigma B\Sigma)\} \quad + \quad 2(\mathrm{tr}V_N^2)\mathrm{tr}(A\Sigma B\Sigma)$$

$$+ 2 \sum_j \sum_k \sum_\ell \sum_m a_{jk}b_{\ell m}\{\mu_3(k,\ell,m)\underset{\sim}{\eta}^{(j)\prime} + \mu_3(m,j,k)\underset{\sim}{\eta}^{(\ell)\prime}\}V_N\underset{\sim}{v}_N$$

$$+ 4\mathrm{tr}(A\Sigma B\P'V_N^2\P)],$$

where V_N is given by (2.6) and $\underset{\sim}{v}_N$ is the column vector of the diagonal elements of V_N.

Proof. Note that $\mathrm{tr}(A\hat{\Sigma}_N) = (\mathrm{tr}V_N)^{-1}\mathrm{tr}(AY'V_NY)$ and V_N is symmetric. Then the results follows from THEOREM A.1 in Appendix.

COROLLARY 4.1. Let $\hat{\sigma}_N(j,\ell)$, $1 \le j,\ \ell \le p$, be the (j,ℓ)-element of $\hat{\Sigma}_N$.
Then, under the same assumptions as in THEOREM 4.1

(4.4) $$\mathrm{Cov}[\hat{\sigma}_N(j,k),\ \hat{\sigma}_N(\ell,m)]$$

$$= (\mathrm{tr}V_N)^{-2}[\underset{\sim}{v}_N'\underset{\sim}{v}_N\{\mu_4(j,k,\ell,m) - \sigma_{jk}\sigma_{\ell m} - \sigma_{j\ell}\sigma_{km} - \sigma_{jm}\sigma_{\ell k}\}$$

$$+ (\mathrm{tr}V_N^2)(\sigma_{j\ell}\sigma_{km} + \sigma_{jm}\sigma_{k\ell})$$

$$+ \mu_3(k,\ell,m)\underset{\sim}{v}_N'V_N\underset{\sim}{\eta}^{(j)} + \mu_3(j,\ell,m)\underset{\sim}{v}_N'V_N\underset{\sim}{\eta}^{(k)}$$

$$+ \mu_3(m,j,k)\underset{\sim}{v}_{\mathcal{N}}'V_{\mathcal{N}}\underset{\sim}{\eta}^{(\ell)} + \mu_3(\ell,j,k)\underset{\sim}{v}_{\mathcal{N}}'V_{\mathcal{N}}\underset{\sim}{\eta}^{(m)}$$

$$+ \sum_{\alpha,\beta\in K}\sum c_\alpha^4 c_\beta^4 u_{\alpha\beta}(\sigma_{j\ell}\xi_{k\alpha}\xi_{m\beta} + \sigma_{jm}\xi_{k\alpha}\xi_{\ell\beta}$$

$$+ \sigma_{k\ell}\xi_{j\alpha}\xi_{m\beta} + \sigma_{km}\xi_{j\alpha}\xi_{\ell\beta})],$$

where $u_{\alpha\beta} = (\underset{\sim}{w}_\alpha - \underset{\sim}{\delta}_\alpha)'(\underset{\sim}{w}_\beta - \underset{\sim}{\delta}_\beta)$.

Proof. The result is obtained from (4.3) by letting $A = (\underset{\sim}{\delta}_j\underset{\sim}{\delta}_k' + \underset{\sim}{\delta}_k\underset{\sim}{\delta}_j')/2$, $B = (\underset{\sim}{\delta}_\ell\underset{\sim}{\delta}_m' + \underset{\sim}{\delta}_m\underset{\sim}{\delta}_\ell')/2$ and $V = (\mathrm{tr}V_{\mathcal{N}})^{-1}V_{\mathcal{N}}$ and using the identities $\underset{\sim}{\eta}^{(\alpha)\prime}V_{\mathcal{N}}\underset{\sim}{\eta}^{(\beta)} = \sum_{i\in K} c_i^4\xi_{i\alpha}\xi_{i\beta}$.

COROLLARY 4.2. If $\underset{\sim}{\varepsilon}_1, \cdots, \underset{\sim}{\varepsilon}_n$ are independently distributed according to $N_p(\underset{\sim}{0}, \Sigma)$, then

$$(4.5) \qquad \mathrm{Cov}[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}}), \mathrm{tr}(B\hat{\Sigma}_{\mathcal{N}})] = 2(\mathrm{tr}V_{\mathcal{N}})^{-2}[\mathrm{tr}V_{\mathcal{N}}^2\ \mathrm{tr}(A\Sigma B\Sigma) + 4\mathrm{tr}(A\Sigma B \P' V_{\mathcal{N}}^2 \P)],$$

for any $p \times p$ symmetric matrices A and B.

Proof. The result is obtained from COROLLARY A.1 by letting $V = (\mathrm{tr}V_{\mathcal{N}})^{-1}V_{\mathcal{N}}$.

It is interesting to compare $\hat{\Sigma}_{\mathcal{L}}$ (or $\hat{\Sigma}_{\mathfrak{U}}$) with the best linear unbiased estimator $\hat{\Sigma}_{BLUE}$ under the linear regression model. Let $V_{\mathcal{L}}$ and $V_{\mathfrak{U}}$ be the matrices obtained from the matrix $V_{\mathcal{N}}$ in (2.6) by using the weight-vectors (2.9) and (2.8), respectively. To compare $\hat{\Sigma}_{\mathcal{L}}$ with $\hat{\Sigma}_{BLUE}$, consider the case when the regression function $\underset{\sim}{\eta}$ is exactly linear, and is given

$$E[Y] = \eta = X\Theta,$$

where Θ is a $q \times p$ matrix of unknown parameters. Letting $P_X = X(X'X)^{-1}X$, the *best linear unbiased* estimator is given by

$$\hat{\Sigma}_{BLUE:\mathcal{L}} = Y'(I_n - P_X)Y/(n-q).$$

As a criterion for the efficiency of $\hat{\Sigma}_{\mathcal{L}}$, we consider the ratio

$$\rho_{\mathcal{L}}(A) = \mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{BLUE:\mathcal{L}})]/\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{L}})],$$

where A is a $p \times p$ symmetric matrix. Note that

$$(4.6) \qquad V_{\mathcal{L}}X = \sum_{i \in K} c_i^4\{D_iX(X'D_iX)^{-}\underset{\sim}{x}_i - \underset{\sim}{\delta}_i\}\{D_iX(X'D_iX)^{-1}\underset{\sim}{x}_i - \underset{\sim}{\delta}_i\}'X$$

$$= 0_{n\times q},$$

and $(I - P_X)X = 0_{n\times q}$. Using these properties and COROLLARY 4.2, we obtain

$$\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{L}})] = 2(\mathrm{tr}V_{\mathcal{L}})^{-2}\mathrm{tr}V_{\mathcal{L}}^2\mathrm{tr}(A\Sigma)^2,$$

$$\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{BLUE:\mathcal{L}})] = 2(n-q)^{-1}\mathrm{tr}(A\Sigma)^2,$$

if $\underset{\sim}{\varepsilon}_i'$s are normally disributed. Thus the ratio $\rho_{\mathcal{L}}(A)$ does not depend on the choice of A in this situation and is given by

$$\rho_{\mathcal{L}} = \{(\mathrm{tr}V_{\mathcal{L}})^2/(\mathrm{tr}V_{\mathcal{L}}^2)\}/(n-q) = \nu_{\mathcal{L}}/(n-q).$$

As for the range of $\rho_{\mathcal{L}}$ we have the following theorem:

THEOREM 4.2. Let $\rho_{\mathcal{L}} = \nu_{\mathcal{L}}/(n-q)$, $\nu_{\mathcal{L}} = (trV_{\mathcal{L}})^2/trV_{\mathcal{L}}^2$, $g_n = \max_{i\in K} \|\underset{\sim}{w}_i\|^2$ and

$$U_n = \max_{\alpha\in K} \#\{\beta \mid N_\beta \cap N_\alpha \neq \phi\}. \tag{4.7}$$

Then

$$(n-q)^{-1}\cdot\max\{\frac{n_K}{(1+g_n)U_n}, 1\} \leq \rho_{\mathcal{L}} \leq \min\{\frac{n_K}{n-q}, 1\}. \tag{4.8}$$

THEOREM 4.2 is a direct consequence of the following lemma:

LEMMA 4.1. Let $\nu_{\mathcal{N}} = (trV_{\mathcal{N}})^2/trV_{\mathcal{N}}^2$, $\nu_{\mathcal{L}} = (trV_{\mathcal{L}})^2/trV_{\mathcal{L}}^2$ and $\nu_{\mathcal{U}} = (trV_{\mathcal{U}})^2/trV_{\mathcal{U}}^2$. Then

(i) $$\max\{\frac{n_K}{(1+g_n)U_n}, 1\} \leq \nu_{\mathcal{N}} \leq n_K,$$

(ii) $$\nu_{\mathcal{L}} \leq n - q,$$

(iii) $$\nu_{\mathcal{U}} \leq n - 1.$$

Proof. Since $c_i^2 = (1 + \|\underset{\sim}{w}_i\|^2)^{-1} < 1$ for $i \in K$, we have

$$\begin{aligned} trV_{\mathcal{N}}^2 &= \sum_\alpha \sum_\beta c_\alpha^4 c_\beta^4 \{(\underset{\sim}{w}_\alpha - \underset{\sim}{\delta}_\alpha)'(\underset{\sim}{w}_\beta - \underset{\sim}{\delta}_\beta)\}^2 \\ &= \sum_\alpha \sum_{\beta:N_\beta\cap N_\alpha\neq\phi} c_\alpha^4 c_\beta^4 \{(\underset{\sim}{w}_\alpha - \underset{\sim}{\delta}_\alpha)'(\underset{\sim}{w}_\beta - \underset{\sim}{\delta}_\beta)\}^2 \\ &\leq \sum_{\alpha\in K} \sum_{\beta:N_\alpha\cap N_\beta\neq\phi} c_\alpha^4 c_\beta^4\cdot(1 + \|\underset{\sim}{w}_\alpha\|^2)(1 + \|\underset{\sim}{w}_\beta\|^2) \end{aligned}$$

$$= \sum_{\alpha \in K} c_\alpha^2 \sum_{\beta : N_\alpha \cap N_\beta \neq \phi} c_\beta^2$$

$$\leq \min\{(\mathrm{tr}V_{\mathcal{N}})^2, \ U_n \mathrm{tr}V_{\mathcal{N}}\}.$$

Therefore, it follows that $\nu_{\mathcal{N}} \geq 1$ and

$$\nu_{\mathcal{N}} \geq (\mathrm{tr}V_{\mathcal{N}})^2/(U_n \mathrm{tr}V_{\mathcal{N}}) = \sum_{i \in K} c_i^2/U_n \geq n_K/\{(1+g_n)U_n\}.$$

The remaining part of (i) is proved from the Cauchy-Schwarz inequality as follows:

$$(\mathrm{tr}V_{\mathcal{N}})^2 = \left(\sum_{i \in K} c_i^2\right)^2 \leq \left(\sum_{i \in K} 1\right)\left(\sum_{i \in K} c_i^4\right) = n_K\left(\sum_{i \in K} c_i^4\right) \leq n_K \mathrm{tr}V_{\mathcal{N}}^2.$$

For the proof of (ii), consider

(4.9) $$W_{\mathcal{L}} = (n-q)^{-1}(I - P_X) - (\mathrm{tr}V_{\mathcal{L}})^{-1}V_{\mathcal{L}}.$$

Since $\mathrm{tr}(P_X V_{\mathcal{L}}) = 0$ yields from (4.6), we have

(4.10) $$\begin{aligned} \mathrm{tr}W_{\mathcal{L}}^2 &= (n-q)^{-2}\mathrm{tr}(I_n - P_X)^2 + (\mathrm{tr}V_{\mathcal{L}})^{-2}\mathrm{tr}V_{\mathcal{L}}^2 \\ &\quad - 2(n-q)^{-1}(\mathrm{tr}V_{\mathcal{L}})^{-1}\mathrm{tr}\{(I_n - P_X)V_{\mathcal{L}}\} \\ &= (n-q)^{-1} + \nu_{\mathcal{L}}^{-1} - 2(n-q)^{-1}(\mathrm{tr}V_{\mathcal{L}})^{-1}\mathrm{tr}V_{\mathcal{L}} \\ &= \nu_{\mathcal{L}}^{-1} - (n-q)^{-1}. \end{aligned}$$

Therefore, noting that $\mathrm{tr}W_{\mathcal{L}}^2 \geq 0$, we obtain $\nu_{\mathcal{L}} \leq n-q$. Similarly (iii) is proved by considering

(4.11) $$W_{\mathfrak{U}} = (n-1)^{-1}(I_n - P_{\underset{\sim}{1}_n}) - (\mathrm{tr}V_{\mathfrak{U}})^{-1}V_{\mathfrak{U}},$$

where $P_{\underset{\sim}{1}_n} = \frac{1}{n}\underset{\sim}{1}_n\underset{\sim}{1}_n'$.

Similarly the efficiency of $\hat{\Sigma}_{\mathfrak{U}}$ may be measured by

$$\rho_{\mathfrak{U}}(A) = \mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathrm{BLUE}:\mathfrak{U}})]/\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathfrak{U}})],$$

where A is a p × p symmetric matrix and $\hat{\Sigma}_{\mathrm{BLUE}:\mathfrak{U}} = (n-1)^{-1}Y'(I_n - P_{\underset{\sim}{1}_n})Y$. It is easily seen that if $\underset{\sim}{\varepsilon}_i$'s are normally distributed, $\rho_{\mathfrak{U}}(A)$ does not depend on A and is given by

$$\rho_{\mathfrak{U}} = \{(\mathrm{tr}V_{\mathfrak{U}})^2/\mathrm{tr}V_{\mathfrak{U}}^2\}/(n-1) = \nu_{\mathfrak{U}}/(n-1).$$

As for the range of $\rho_{\mathfrak{U}}$, we have the following theorem:

THEOREM 4.3. Let $\rho_{\mathfrak{U}} = \nu_{\mathfrak{U}}/(n-1)$ and $\nu_{\mathfrak{U}} = (\mathrm{tr}V_{\mathfrak{U}})^2/\mathrm{tr}V_{\mathfrak{U}}^2$. Then

(4.12) $$(n-1)^{-1}\cdot\max\{\frac{n_K}{2U_n}, 1\} \le \rho_{\mathfrak{U}} \le \min\{\frac{n_K}{n-1}, 1\}.$$

Proof. The results follows from LEMMA 4.1 and $n_i\|\underset{\sim}{w}_i\|^2 = 1$ $(i \in K)$ for $\hat{\Sigma}_{\mathfrak{U}}$.

5. Asymptotic properties

It is easily expected that the asymptotic behaviors of $\hat{\Sigma}_{\mathfrak{N}}$ depend sensitively on the design of the explanatory variables as well as on the error distribution. We first postulate the following conditions on them.

CONDITION 1. $\nu_{\mathcal{N}} = (trV_{\mathcal{N}})^2/trV_{\mathcal{N}}^2 \to +\infty$, as $n \to +\infty$.

CONDITION 2. There exists a positive number G such that

$$\max_{i \in K} n_i \|\underset{\sim}{w}_i\|^2 \le G < +\infty.$$

CONDITION 3. The errors $\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2, \cdots$ are independently distributed with finite fourth moments.

REMARK 5.1. CONDITION 2 is fulfilled for a LUW estimator, since $n_i \|\underset{\sim}{w}_i\|^2 = 1$ for all $i \in K$ in this case.

In this section the eigenvalues of several symmetric matrices will be frequently operated; for simplicity, we shall express the jth largest eigenvalue of a symmetric matrix A as $\lambda_j(A)$.

We now prove the consistency of $\hat{\Sigma}_{\mathcal{N}}$ which is given in the following theorem:

THEOREM 5.1. Suppose that CONDITIONS 1, 2 and 3 hold. Then, the nonparametric estimator $\hat{\Sigma}_{\mathcal{N}}$ of (2.2) is consistent if

(5.1) $$\sum_{i \in K} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i = o(n_K), \quad \text{as } n \to +\infty,$$

where $\underset{\sim}{\xi}_i = E[\underset{\sim}{r}_i]$ for $i \in K$.

Proof. It is sufficient to show that $tr(A\hat{\Sigma}) \to tr(A\Sigma)$ as $n \to +\infty$ in probability, for any symmetric $p \times p$ matrix A. First we show that

$E[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}})] \to E[\mathrm{tr}(A\Sigma)]$ as $n \to +\infty$. Since $|\underset{\sim}{\xi}_i' A \underset{\sim}{\xi}_i| \le \max_j |\lambda_j(A)| \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i$ and $1/(1+G) \le c_i^2 < 1$ for any $i \in K$, we obtain from (A.2) in Appendix that

$$|E[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}})] - \mathrm{tr}(A\Sigma)| = |(\mathrm{tr}V_{\mathcal{N}})^{-1}E[\mathrm{tr}(AY'V_{\mathcal{N}}Y)] - \mathrm{tr}(A\Sigma)|$$

$$= |(\mathrm{tr}V_{\mathcal{N}})^{-1}\{(\mathrm{tr}V_{\mathcal{N}})\mathrm{tr}(A\Sigma) + \mathrm{tr}(A\P' V_{\mathcal{N}}\P)\} - \mathrm{tr}(A\Sigma)|$$

$$= (\mathrm{tr}V_{\mathcal{N}})^{-1}|\mathrm{tr}(A\P' V_{\mathcal{N}}\P)|$$

$$= \Big(\sum_{i\in K} c_i^2\Big)^{-1} \Big|\sum_{i\in K} c_i^4\, \underset{\sim}{\xi}_i' A \underset{\sim}{\xi}_i\Big|$$

$$\le \max_j |\lambda_j(A)|(1+G)\Big(\sum_i \underset{\sim}{\xi}_i'\underset{\sim}{\xi}_i/n_K\Big).$$

Thus, it follows from (5.1) that $E[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}})] \to \mathrm{tr}(A\Sigma)$ as $n \to +\infty$.

Next we show that $\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}})] \to 0$ as $n \to +\infty$. Since $c_i^2 < 1$ and $\underset{\sim}{v}_{\mathcal{N}}'\underset{\sim}{v}_{\mathcal{N}} \le \mathrm{tr}V_{\mathcal{N}}^2$,

$$|\underset{\sim}{\eta}^{(j)\prime} V_{\mathcal{N}}\underset{\sim}{v}_{\mathcal{N}}| \le \Big(\underset{\sim}{\eta}^{(j)\prime} V_{\mathcal{N}}\underset{\sim}{\eta}^{(j)}\underset{\sim}{v}_{\mathcal{N}}' V_{\mathcal{N}}\underset{\sim}{v}_{\mathcal{N}}\Big)^{\frac{1}{2}}$$

$$\le \Big(\sum_{i\in K} c_i^4\xi_{ij}^2\Big)^{\frac{1}{2}}\{\lambda_1(V_{\mathcal{N}})\underset{\sim}{v}_{\mathcal{N}}'\underset{\sim}{v}_{\mathcal{N}}\}^{\frac{1}{2}}$$

$$\le \Big(\sum_{i\in K} \underset{\sim}{\xi}_i'\underset{\sim}{\xi}_i\Big)^{\frac{1}{2}}\{\lambda_1(V_{\mathcal{N}})(\mathrm{tr}V_{\mathcal{N}}^2)\}^{\frac{1}{2}}.$$

Letting $A = [\underset{\sim}{a}_1, \cdots, \underset{\sim}{a}_p]$,

(5.2) $$|\mathrm{tr}(A\Sigma A\P' V_{\mathcal{N}}^2\P)| \le \sum_j\sum_k |\underset{\sim}{a}_j'\Sigma\underset{\sim}{a}_k\underset{\sim}{\eta}^{(j)\prime}V_{\mathcal{N}}^2\underset{\sim}{\eta}^{(k)}|$$

$$\le \sum_j\sum_k |\underset{\sim}{a}_j'\Sigma\underset{\sim}{a}_k|\Big(\underset{\sim}{\eta}^{(j)\prime}V_{\mathcal{N}}^2\underset{\sim}{\eta}^{(j)}\underset{\sim}{\eta}^{(k)\prime}V_{\mathcal{N}}^2\underset{\sim}{\eta}^{(k)}\Big)^{\frac{1}{2}}$$

$$\leq \lambda_1(V_{\mathcal{N}})\left(\sum_{i\in K}\xi_i'\xi_i\right)\sum_j\sum_k |a_j'\Sigma a_k|.$$

Therefore, using (4.3) we obtain

$$\mathrm{var}[\mathrm{tr}(A\hat{\Sigma})] \leq \nu_{\mathcal{N}}^{-1}[|\sum_j\sum_k\sum_\ell\sum_m a_{jk}a_{\ell m}\mu_4(j,k,\ell,m) - \{\mathrm{tr}(A\Sigma)\}^2$$

$$- 2\mathrm{tr}(A\Sigma)^2]| + 2\mathrm{tr}(A\Sigma)^2]$$

$$+ 2\sqrt{(1+G)/\nu_{\mathcal{N}}}\ \{\lambda_1(V_{\mathcal{N}})/\mathrm{tr}V_{\mathcal{N}}\}^{\frac{1}{2}}\left(\sum_{i\in K}\xi_i'\xi_i/n_K\right)^{\frac{1}{2}}$$

$$\times \sum_j\sum_k\sum_\ell\sum_m |a_{jk}a_{\ell m}|\{|\mu_3(k,\ell,m)| + |\mu_3(m,j,k)|\}$$

$$+ 4(1+G)^2\left(\sum_{i\in K}\xi_i'\xi_i/n_K\right)\{\lambda_1(V_{\mathcal{N}})/n_K\}\sum_j\sum_\ell |a_j'\Sigma a_\ell|.$$

This implies that under CONDITIONS 1, 2 and 3, $\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}})] \to 0$ as $n \to +\infty$.

Finally, using the Chebychev's inequality, we obtain that for any $\varepsilon > 0$

$$\Pr\{|\mathrm{tr}\{A(\hat{\Sigma}_{\mathcal{N}} - \Sigma)\} \geq \varepsilon\} \leq \frac{\mathrm{Var}[\mathrm{tr}(A\hat{\Sigma}_{\mathcal{N}})] + \{E[\mathrm{tr}\{A(\hat{\Sigma}_{\mathcal{N}} - \Sigma)\}]\}^2}{\varepsilon^2} \to 0,$$

as $n \to +\infty$. This completes the proof.

COROLLARY 5.1. Suppose that CONDITIONS 1 and 3 hold and that the regression function $\underset{\sim}{\eta}$ is differentiable and satisfies

(5.3) $$\psi_\alpha = \sup_{\underset{\sim}{t}}\left\{\sum_{\ell=1}^{q}\left(\frac{\partial}{\partial x_\ell}\eta_\alpha(\underset{\sim}{x})|_{\underset{\sim}{x}=\underset{\sim}{t}}\right)^2\right\}^{\frac{1}{2}} < +\infty, \quad \alpha = 1, \cdots, p.$$

Then a LUW estimator $\hat{\Sigma}_{\mathcal{U}}$ of Σ is consistent if

$$(5.4) \qquad \sum_{i \in K} d_i^2 = o(n_K) \qquad \text{as } n \to +\infty.$$

Proof. For a LUW estimator, CONDITION 2 is automatically fulfilled (see, REMARK 5.1), and it yields from LEMMA 3.1 that under assumptions (5.3) and (5.4)

$$0 \leq \sum_{i \in K} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i / n_K \leq \left(\sum_{\alpha=1}^{p} \psi_\alpha^2 \right) \left(\sum_{i \in K} d_i^2 / n_K \right) \to 0,$$

as $n \to +\infty$. Hence, the assertion follows from THEOREM 5.1.

COROLLARY 5.2. Suppose that CONDITIONS 1, 2 and 3 hold and that the regression function $\underset{\sim}{\eta}$ is twice differentiable and has the Hessian satisfying

$$(5.5) \qquad \gamma_\alpha = \sup_{\underset{\sim}{x}} \sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} |\underset{\sim}{u}' H_{\underset{\sim}{x}}^{(\alpha)} \underset{\sim}{u}| < +\infty, \qquad \alpha = 1, \cdots, p.$$

Then a LLW estimator $\hat{\Sigma}_{\mathcal{L}}$ of Σ is consistent if

$$(5.6) \qquad \sum_{i \in K} d_i^4 = o(n_K), \qquad \text{as } n \to +\infty.$$

Proof. For a LLW estimator, it yields from LEMMA 3.2 that under the assumption (5.5) and (5.6)

$$0 \leq \sum_{i \in K} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i / n_K \leq \frac{1}{4} G \left(\sum_{\alpha=1}^{p} \gamma_\alpha^2 \right) \left(\sum_{i \in K} d_i^4 / n_K \right) \to 0,$$

as $n \to +\infty$. Hence, the assertion follows from THEOREM 5.1.

To derive the asymptotic normality of $\hat{\Sigma}_{\mathcal{N}}$, somewhat stronger conditions are needed on the error distribution and on the design; we now postulate the following conditions:

CONDITION 3†. The errors $\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2, \cdots$ are independently distributed according to $N_p(\underset{\sim}{0}, \Sigma)$.

CONDITION 4. $\lambda_1(V_N) = o(\sqrt{n_K})$, as $n \to +\infty$.

THEOREM 5.2. Suppose that CONDITIONS 1, 2, 3† and 4 hold. If

$$\sum_{i \in K} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i = o(\sqrt{n_K}), \qquad \text{as } n \to +\infty, \tag{5.7}$$

then the asymptotic distribution of $Z_N = \sqrt{\nu_N}(\hat{\Sigma}_N - \Sigma)$ is normal with mean $\underset{\sim}{0}$ and covariances

$$E[z_{jk} z_{\ell m}] = \sigma_{j\ell}\sigma_{km} + \sigma_{jm}\sigma_{k\ell}. \tag{5.8}$$

Proof. It is sufficient to show that every linear function of Z_N has an asymptotic univariate normal distibution (see, e.g., Rao[15, Chapter 8a.2]). Note that an arbitrary linear combination of Z_N can be written as

$$tr(AZ_N) = \sqrt{\nu_N} tr[A(\hat{\Sigma}_N - \Sigma)],$$

where A be a symmetric p × p matrix. A few algebra yields that the quantity $tr(AZ_N)$ can be decomposed into the following three terms:

$$tr(AZ_N) = tr(AZ_N^*) + \xi_N(A) + \tau_N(A),$$

where

$$Z_{\mathcal{N}}^{*} = \sqrt{\nu_{\mathcal{N}}}\{(\mathrm{tr}V_{\mathcal{N}})^{-1}\mathcal{E}'V_{\mathcal{N}}\mathcal{E} - \Sigma\},$$

$$\xi_{\mathcal{N}}(A) = \sqrt{\nu_{\mathcal{N}}}\ \mathrm{tr}[A\P'V_{\mathcal{N}}\P]/\mathrm{tr}V_{\mathcal{N}},$$

$$\varphi_{\mathcal{N}}(A) = 2\sqrt{\nu_{\mathcal{N}}}\ \mathrm{tr}[A\P'V_{\mathcal{N}}\mathcal{E}]/\mathrm{tr}V_{\mathcal{N}}.$$

Then under CONDITIONS 1, 2 and (5.7) we obtain

$$\begin{aligned}|\xi_{\mathcal{N}}(A)| &= \sqrt{\nu_{\mathcal{N}}}|\mathrm{tr}[A\sum_{i\in K} c_i^4 \underset{\sim}{\xi}_i \underset{\sim}{\xi}_i']|/\mathrm{tr}V_{\mathcal{N}} \\ &\le \sqrt{n_K}\ \sum_{i\in K} c_i^4 |\underset{\sim}{\xi}_i' A \underset{\sim}{\xi}_i| / \sum_{i\in K} c_i^2 \\ &\le \max_j |\lambda_j(A)|(1+G)\sum_{i\in K}\underset{\sim}{\xi}_i'\underset{\sim}{\xi}_i/\sqrt{n_K} \\ &\to 0,\end{aligned}$$

as $n \to +\infty$. Since $\varphi_{\mathcal{N}}(A)$ is a linear combination of $\underset{\sim}{\varepsilon}_i$'s, $\varphi_{\mathcal{N}}(A)$ is normally distrbuted with mean 0 and variance

$$\mathrm{Var}[\varphi_{\mathcal{N}}(A)] = 4\nu_{\mathcal{N}}(\mathrm{tr}V_{\mathcal{N}})^{-2}E[\{\mathrm{tr}(A\P'V_{\mathcal{N}}\mathcal{E})\}^2].$$

Since $(1+G)^{-1} \le c_i^2 < 1$ $(i \in K)$, it follows from (A.5) and (5.2) that

$$\begin{aligned}E[\{\mathrm{tr}(A\P'V_{\mathcal{N}}\mathcal{E})\}^2] &= \mathrm{tr}(A\Sigma A\P'V_{\mathcal{N}}^2\P) \\ &\le \left(\sum_j \underset{\sim}{a}_j'\Sigma\underset{\sim}{a}_j\right)\lambda_1(V_{\mathcal{N}})\sum_{i\in K}\underset{\sim}{\xi}_i'\underset{\sim}{\xi}_i.\end{aligned}$$

Hence, under CONDITIONS 1, 2, 4 and (5.7) we obtain

$$\mathrm{Var}[\varphi_{\mathcal{N}}(A)] \le 4(1+G)^2\left(\sum_j \underset{\sim}{a}_j'\Sigma\underset{\sim}{a}_j\right)\left(\nu_{\mathcal{N}}/n_K\right)\left(\lambda_1(V_{\mathcal{N}})/\sqrt{n_K}\right)\left(\sum_{i\in K}\underset{\sim}{\xi}_i'\underset{\sim}{\xi}_i/\sqrt{n_K}\right) \to 0,$$

as $n \to +\infty$. This implies that $\varphi_{\mathcal{N}}(A) \to 0$ as $n \to +\infty$ in probability.

Next we show that the asymptotic distribution of $\mathrm{tr}(AZ_{\mathcal{N}}^*)$ is normal with mean 0 and variance $2\mathrm{tr}(A\Sigma)^2$. Let $\phi_A(t)$ be the characteristic function of $\mathrm{tr}(AZ_{\mathcal{N}}^*)$. Then

$$(5.9) \qquad \phi_A(t) = E[\exp\{it\cdot\mathrm{tr}(AZ_{\mathcal{N}}^*)\}]$$

$$= E[\exp\{it(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{1}{2}}\mathrm{tr}[A(\mathcal{E}'V_{\mathcal{N}}\mathcal{E}) - (\mathrm{tr}V_{\mathcal{N}})\mathrm{tr}(A\Sigma)]\}].$$

Using an orthogonal transformation of $V_{\mathcal{N}}$, we have

$$(5.10) \qquad \mathrm{tr}[A(\mathcal{E}'V_{\mathcal{N}}\mathcal{E})] = \sum_{\alpha=1}^{n}\lambda_\alpha(V_{\mathcal{N}})\underset{\sim}{\varepsilon}_\alpha^{*\prime}A\underset{\sim}{\varepsilon}_\alpha^*,$$

where $\underset{\sim}{\varepsilon}_\alpha^*$'s are independently distributed according to $N_p(\underset{\sim}{0}, \Sigma)$. Considering the transformation $\underset{\sim}{u}_\alpha = \Sigma^{-\frac{1}{2}}\underset{\sim}{\varepsilon}_\alpha^*$, $\alpha = 1, \cdots, n$, we can write

$$(5.11) \qquad \underset{\sim}{\varepsilon}_\alpha^{*\prime}A\underset{\sim}{\varepsilon}_\alpha^* = \sum_{j=1}^{p}\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})u_{\alpha j}^2,$$

where $u_{\alpha j}$'s are independently distributed according to $N(0,1)$. Hence, from (5.9), (5.10) and (5.11) we obtain that

$$\phi_A(t) = E[\exp\{it(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{1}{2}}\sum_\alpha\lambda_\alpha(V_{\mathcal{N}})\sum_j\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})u_{\alpha j}^2$$

$$- it\sqrt{\nu_{\mathcal{N}}}\mathrm{tr}(A\Sigma)\}]$$

$$= E[\exp\{-it\sqrt{\nu_{\mathcal{N}}}\mathrm{tr}(A\Sigma)\}$$

$$\times \prod_{\alpha} \prod_{j} E[\exp\{it(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{1}{2}}\lambda_\alpha(V_{\mathcal{N}})\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})u_{\alpha j}^2\}].$$

$$= \exp\{-it\sqrt{\nu_{\mathcal{N}}}\mathrm{tr}(A\Sigma)\}$$

$$\times \prod_{\alpha} \prod_{j} \{1 - 2it\cdot(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{1}{2}}\lambda_\alpha(V_{\mathcal{N}})\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})\}^{-\frac{1}{2}}.$$

Note that $\sum_{\alpha} \lambda_\alpha(V_{\mathcal{N}}) = \mathrm{tr}V_{\mathcal{N}}$, $\sum_{\alpha} \{\lambda_\alpha(V_{\mathcal{N}})\}^2 = \mathrm{tr}V_{\mathcal{N}}^2$ and $\sum_{j} \{\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})\}^2 = \mathrm{tr}(A\Sigma)^2$. Then, using a Taylor expansion of $\log \phi_A(t)$, we obtain that for any $t \in (-\infty, +\infty)$

$$\log \phi_A(t) = -t^2\mathrm{tr}(A\Sigma)^2 + \sum_{\alpha}\sum_{j} R_{\alpha j}(t),$$

where

$$R_{\alpha j}(t) = -\frac{4}{3}\, i\, \frac{(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{3}{2}}\{\lambda_\alpha(V_{\mathcal{N}})\}^3\{\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})\}^3}{\left\{1 - 2i\theta_t t(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{1}{2}}\lambda_\alpha(V_{\mathcal{N}})\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})\right\}^3}\, t^3,$$

for some θ_t in $(0, 1)$. Since

$$|R_{\alpha j}(t)| \le \frac{4}{3}t^3(\mathrm{tr}V_{\mathcal{N}}^2)^{-\frac{3}{2}}\lambda_1(V_{\mathcal{N}})\cdot\max_{\ell}|\lambda_\ell(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})|\{\lambda_\alpha(V_{\mathcal{N}})\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})\}^2,$$

it follows that under CONDITIONS 1,2 and 4

$$|\sum_{\alpha}\sum_{j} R_{\alpha j}(t)| \le \sum_{\alpha}\sum_{j} |R_{\alpha j}(t)|$$

$$\le \frac{4}{3}t^3(1+G)\{\lambda_1(V_{\mathcal{N}})/\sqrt{n_K}\}\cdot\max_{j}|\lambda_j(\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})|\mathrm{tr}(A\Sigma)^2 \to 0,$$

as $n \to +\infty$. Therefore, we obtain that $\phi_A(t) \to \exp[-\mathrm{tr}(A\Sigma)^2t^2]$ as $n \to +\infty$ for any $t \in (-\infty, +\infty)$. This completes the proof.

Using the similar argument as in the proof of COROLLARY 5.1 or 5.2, we obtain from THEOREM 5.2 the following corollaries:

COROLLARY 5.3. Suppose that CONDITIONS 1, $3^{\dagger}$ and 4 hold and that the regression function η is differentiable and satisfies (5.3). Then, the asymptotic distribution of $Z_{\mathcal{U}} = \sqrt{\nu_{\mathcal{U}}}(\hat{\Sigma}_{\mathcal{U}} - \Sigma)$ is normal with mean 0 and covariances given by (5.8) if $\sum_{i \in K} d_i^2 = o(\sqrt{n_K})$ as $n \to +\infty$.

COROLLARY 5.4. Suppose that CONDITIONS 1, 2, $3^{\dagger}$ and 4 hold and that the regression function η is twice differentiable and satisfies (5.5). Then, the asymptotic distribution of $Z_{\mathcal{L}} = \sqrt{\nu_{\mathcal{L}}}(\hat{\Sigma}_{\mathcal{L}} - \Sigma)$ is normal with mean 0 and covariances given by (5.8) if $\sum_{i \in K} d_i^4 = o(\sqrt{n_K})$ as $n \to +\infty$.

6. Some special cases when q = 2

We will now consider in more detail the case when $q = 2$. The data may be described as $\{(y_i, x_i) \mid 1 \le i \le n\}$ with $x_i = (1, x_i)'$. Without loss of generality we assume that $x_1 \le x_2 \le \cdots \le x_n$ and the number of repeated observations at x_i is m_i, i.e., $m_i = \#\{j \mid x_j = x_i, 1 \le j \le n\}$. For simplicity, we denote the observations by (y_i, x_i) instead of (y_i, x_i) henceforth. Let $K_{\mathcal{H}} = \{i \mid m_i \ge 2\}$, $K_{\mathcal{G}} = \{2, \cdots, n-1\} \sim K_{\mathcal{H}}$. First we define a practical index set N_i for each $i \in K = K_{\mathcal{H}} \cup K_{\mathcal{G}}$ which specifies a neighborhood of x_i. Let

$$(6.1)\qquad N_i = \begin{cases} \{j \mid x_j = x_i,\ j \neq i\}, & \text{if } i \in K_{\mathscr{N}}, \\ \{j \mid x_j = x_{i^-} \text{ or } x_j = x_{i^+}\}, & \text{if } i \in K_{\mathscr{S}}, \end{cases}$$

where $i^- = \max_{x_\ell < x_i} \ell$ and $i^+ = \min_{x_\ell > x_i} \ell$. For $i \in K_{\mathscr{S}}$, let $N_i^- = \{j \mid x_j = x_{i^-}\}$, $N_i^+ = \{j \mid x_j = x_{i^+}\}$, $m_{i^-} = \#N_i^-$ and $m_{i^+} = \#N_i^+$. It is possible to consider a general estimator $\hat{\Sigma}_{\mathscr{N}}$ of Σ based on N_i, $i \in K$. However, it is natural to consider a simple class of estimators which reflects on the characteristic of two types of neighborhoods as follows: For $i \in K = K_{\mathscr{N}} \cup K_{\mathscr{S}}$ and given $\theta_i \in [0, 1]$, let

$$(6.2)\qquad \hat{\underset{\sim}{y}}_i = (\underset{\sim}{y}^{(1)}, \cdots, \underset{\sim}{y}^{(p)})' \underset{\sim}{w}_i$$

$$= \begin{cases} \{m_i \bar{\underset{\sim}{y}}_{i\cdot} - \underset{\sim}{y}_i\}/(m_i - 1), & \text{if } i \in K_{\mathscr{N}}, \\ \theta_i \bar{\underset{\sim}{y}}_{i^-\cdot} + (1 - \theta_i) \bar{\underset{\sim}{y}}_{i^+\cdot}, & \text{if } i \in K_{\mathscr{S}}, \end{cases}$$

where for $i \in K_{\mathscr{N}}$, $\bar{\underset{\sim}{y}}_{i\cdot} = m_i^{-1}\left(\underset{\sim}{y}_i + \sum_{j \in N_i} \underset{\sim}{y}_j\right)$ and for $i \in K_{\mathscr{S}}$, $\bar{\underset{\sim}{y}}_{i^-} = m_{i^-}^{-1} \sum_{j \in N_i^-} \underset{\sim}{y}_j$ and $\bar{\underset{\sim}{y}}_{i^+} = m_{i^+}^{-1} \sum_{j \in N_i^+} \underset{\sim}{y}_j$.

Using $\underset{\sim}{r}_i = \hat{\underset{\sim}{y}}_i - \underset{\sim}{y}_i$ as in (2.3), we define a class of estimators of Σ by

$$(6.3)\qquad \hat{\Sigma}_G = \left(\sum_{i \in K} c_i^2\right)^{-1} \sum_{i \in K} c_i^4 \underset{\sim}{r}_i \underset{\sim}{r}_i',$$

where $c_i^2 = (1 + \underset{\sim}{w}_i' \underset{\sim}{w}_i)^{-1}$ and is given by

$$(6.4) \qquad c_i^2 = \begin{cases} 1 - m_i^{-1}, & \text{if } i \in K_{\mathscr{K}}, \\ \{1 + \dfrac{\theta_i^2}{m_{i-}} + \dfrac{(1-\theta_i)^2}{m_{i+}}\}^{-1}, & \text{if } i \in K_{\mathscr{G}}. \end{cases}$$

A special case of this estimator were introduced by Gasser et al. [8] and by Ohtaki[14], and a slightly different estimator was proposed by Rice[16] in univariate regression model. A simple algebra yields that

$$(6.5) \qquad \hat{\Sigma}_G = \tau_{\mathscr{K}} \hat{\Sigma}_{PE} + (1-\tau_{\mathscr{K}}) \hat{\Sigma}_{\mathscr{G}},$$

where $\tau_{\mathscr{K}} = \left(\sum_{i \in K_{\mathscr{K}}} c_i^2\right) / \left(\sum_{i \in K} c_i^2\right)$, $\hat{\Sigma}_{PE}$ is the PEMS estimator (1.3) based on the data $\{(\underset{\sim}{y}_i, \underset{\sim}{x}_i) \mid i \in K_{\mathscr{K}}\}$, and

$$\hat{\Sigma}_{\mathscr{G}} = \left(\sum_{i \in K_{\mathscr{G}}} c_i^2\right)^{-1} \sum_{i \in K_{\mathscr{G}}} c_i^4 \underset{\sim}{r}_i \underset{\sim}{r}_i'.$$

Thus, we can see that $\hat{\Sigma}_G$ is a natural extension of PEMS estimator. Note that $\hat{\Sigma}_G$ includes two important estimators as special cases; adopting $\theta_i = m_{i-}/(m_{i-} + m_{i+})$ yields a LUW estimator, which will be denoted by $\hat{\Sigma}_{G\mathscr{U}}$, and adopting $\theta_i = (x_i - x_{i-})/(x_{i+} - x_{i-})$ yields a LLW estimator, which will be denoted by $\hat{\Sigma}_{G\mathscr{L}}$.

REMARK 6.1. The estimator $\hat{\Sigma}_G$ is expressed in a quadratic form, $\hat{\Sigma}_G = (\mathrm{tr}V_G)^{-1} Y' V_G Y$, where the (α, β)-element of V_G can be expressed as follows:

$$
(6.6) \qquad v_{\alpha\beta}
$$

$$
= \begin{cases}
c_\alpha^4 + I_{\mathscr{S}}(\alpha^-)c_{\alpha^-}^4(1-\theta_{\alpha^-})^2 + I_{\mathscr{S}}(\alpha^+)c_{\alpha^+}^4\theta_{\alpha^+}^2, & \text{if } \alpha = \beta \in K_{\mathscr{S}}, \\
(m_\alpha - 1)/m_\alpha, & \text{if } \alpha = \beta \in K_{\mathscr{N}}, \\
-m_\alpha^{-1} + m_\alpha^{-2}\{I_{\mathscr{S}}(\alpha^-)c_{\alpha^-}^4(1-\theta_{\alpha^-})^2 + I_{\mathscr{S}}(\alpha^+)c_{\alpha^+}^4\theta_{\alpha^+}^2\}, & \\
 & \text{if } x_\alpha = x_\beta, \text{ and } \alpha \neq \beta, \\
-I_{\mathscr{S}}(\alpha)c_\alpha^4\,(1-\theta_\alpha)m_{\alpha^+}^{-1} - I_{\mathscr{S}}(\alpha^+)c_{\alpha^+}^4\theta_{\alpha^+}m_\alpha^{-1}, & \text{if } \beta = \alpha^+, \\
I_{\mathscr{S}}(\alpha^+)c_{\alpha^+}^4\theta_{\alpha^+}(1-\theta_{\alpha^+})m_\alpha^{-1}m_{\alpha^{++}}^{-1}, & \text{if } \beta = \alpha^{++}, \\
0, & \text{otherwise},
\end{cases}
$$

where for $\mathscr{W} = \mathscr{N}, \mathscr{S}$, $I_{\mathscr{W}} = 1$ if $i \in K_{\mathscr{W}}$ and 0 if $i \notin K_{\mathscr{W}}$.

Since $\hat{\Sigma}_G$ (or $\Sigma_{G\mathfrak{U}}$, $\hat{\Sigma}_{G\mathscr{L}}$) is a special one of $\hat{\Sigma}_{\mathscr{N}}$ (or $\hat{\Sigma}_{\mathfrak{U}}$, $\hat{\Sigma}_{\mathscr{L}}$), we can apply the general theory of $\hat{\Sigma}_{\mathscr{N}}$ in Sections 3 ~ 5 to $\hat{\Sigma}_G$ (or $\Sigma_{G\mathfrak{U}}$, $\hat{\Sigma}_{G\mathscr{L}}$). However, $\hat{\Sigma}_G$ is based on a special index-set N_i and a special predictor $\hat{\underset{\sim}{y}}_i$, and so we can expect that the assertions and the conditions in the general theory of $\hat{\Sigma}_{\mathscr{N}}$ can be more strengthened and simplified. We shall look these in the following.

LEMMA 6.1. Let $\hat{\Sigma}_{G\mathfrak{U}} = [\sigma_{G\mathfrak{U}}(j,\ell)]$ be a LUW estimator. Suppose that the jth and the ℓth components η_j and η_ℓ of the regression function $\underset{\sim}{\eta}$ are differentiable, and that

(6.7) $$\psi_\alpha = \sup_t \left|\frac{d}{dx}\eta_\alpha(x)\Big|_{x=t}\right| < +\infty, \qquad \alpha = j,\ \ell.$$

Then $$|E[\hat{\sigma}_{G\mathcal{U}}(j,\ell)] - \sigma_{j\ell}| \le \psi_j\psi_\ell h_{G\mathcal{U}},$$

where

(6.8) $$h_{G\mathcal{U}} = \Big(\sum_{i\in K} c_i^2\Big)^{-1} \sum_{i\in K_{\mathcal{G}}} c_i^4(x_{i+} - x_{i-})^2.$$

LEMMA 6.2. Let $\hat{\Sigma}_{G\mathcal{L}} = [\hat{\sigma}_{G\mathcal{L}}(j,\ell)]$ be a LLW estimator. Suppose that the jth and the ℓth components η_j and η_ℓ of the regression function $\tilde{\eta}$ are twice differentiable, and that

(6.9) $$\gamma_\alpha = \sup_t \left|\frac{d^2}{dx^2}\eta_\alpha(x)\Big|_{x=t}\right| < +\infty, \qquad \alpha = j,\ \ell.$$

Then $$|E[\hat{\sigma}_{G\mathcal{L}}(j,\ell)] - \sigma_{j\ell}| \le \gamma_j\gamma_\ell h_{G\mathcal{L}},$$

where

(6.10) $$h_{G\mathcal{L}} = \frac{1}{4}\Big(\sum_{i\in K} c_i^2\Big)^{-1} \sum_{i\in K_{\mathcal{G}}} c_i^4(x_{i+} - x_i)^2(x_i - x_{i-})^2.$$

LEMMA 6.3. Let V_G be the matrix given in REMARK 6.1 and let $\nu_G = (\mathrm{tr}V_G)^2/\mathrm{tr}V_G^2$. Then it holds that

(6.11) $$\nu_G \ge \frac{n-2}{24 + 40(n-2)^{-1}}.$$

Proof. Note that

(6.12) $$\nu_G = \Big(\sum_{i\in K} c_i^2\Big)^2 / [\sum_{i\in K} c_i^4 \{1 + 2I_{\mathcal{G}}(i^+)c_{i^+}^4(1-\theta_i+\theta_{i^+})^2 + 2I_{\mathcal{G}}(i^{++})c_{i^{++}}^4 \frac{\theta_{i^{++}}^2(1-\theta_i)^2}{m_{i+1}^2}\} + \sum_{i\in K_{\mathcal{H}}} c_i^2].$$

Since $\frac{1}{2} \le c_i^2 < 1$ for all $i \in K$, a straightforward calculation yields that

$$\nu_G \ge \Big(\sum_{i\in K} c_i^2\Big)^2 / \Big(2\sum_{i\in K} c_i^2 + 5n\Big) = \Big(\sum_{i\in K} c_i^2\Big) / \Big(2 + \frac{5n}{\sum_{i\in K} c_i^2}\Big) \ge \frac{n-2}{24 + \frac{40}{n-2}}.$$

Hence we obtain the desired result.

From LEMMA 6.3 it follows that $\nu_G \to +\infty$ as $n \to +\infty$, and CONDITION 1 in Section 5 are satisfied; therefore, we obtain from THEOREM 5.1 the following theorems:

THEOREM 6.1. Suppose that CONDITIONS 2 and 3 in Section 5 hold. If $\sum_{i\in K} \xi_i'\xi_i = o(n)$ as $n \to +\infty$, then the nonparametric estimator $\hat{\Sigma}_G$ deined by (6.3) is consistent.

COROLLARY 6.1. Suppose that CONDITION 3 holds, and that the regression function η is differentiable and satisfies $\sum_{\alpha} \psi_\alpha^2 < +\infty$, where ψ_α's are the quantities given by (6.7). Then a LUW estimator $\hat{\Sigma}_{G\mathfrak{U}}$ is consistent if

$$\sum_{i\in K_{\mathcal{G}}} (x_{i^+} - x_{i^-})^2 = o(n), \qquad \text{as } n \to +\infty.$$

Proof. Using $\frac{1}{2} \le c_i^2 < 1$ $(i \in K)$, we obtain from LEMMA 6.1 that

$$(6.13) \qquad 0 \le \sum_{i \in K} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i \le 2\left(\sum_{\alpha=1}^{p} \psi_\alpha^2\right) \sum_{i \in K_{\mathscr{G}}} (x_{i^+} - x_{i^-})^2.$$

Hence, the assertion follows.

COROLLARY 6.2. Suppose that CONDITION 3 holds, and that the regression function $\underset{\sim}{\eta}$ is twice differentiable and satisfies $\sum_\alpha \gamma_\alpha^2 < +\infty$, where γ_α's are the quantities given by (6.9). Then a LLW estimator $\hat{\Sigma}_{G\mathscr{L}}$ is consistent if

$$\sum_{i \in K_{\mathscr{G}}} (x_{i^+} - x_i)^2 (x_i - x_{i^-})^2 = o(n), \qquad \text{as } n \to +\infty.$$

Proof. Using a similar argument as in the proof of COROLLARY 6.2, we obtain from LEMMA 6.2 that

$$(6.14) \qquad 0 \le \sum_{i \in K} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i \le \frac{1}{2}\left(\sum_{\alpha=1}^{p} \gamma_\alpha^2\right) \sum_{i \in K_{\mathscr{G}}} (x_{i^+} - x_i)^2 (x_i - x_{i^-})^2.$$

Hence, the assertion follows.

COROLLARY 6.3. Suppose that CONDITION 3 holds, and that there exist two numbers a and b such that $-\infty < a \le x_i \le b < +\infty$ for all $i \in K$. Then, $\hat{\Sigma}_{G\mathfrak{U}}$ is consistent if the regression function $\underset{\sim}{\eta}$ is differentiable on [a, b]; so is also $\hat{\Sigma}_{G\mathscr{L}}$ if $\underset{\sim}{\eta}$ is twice differentiable on [a, b].

Proof. Let t_j $(j \in K_{\mathscr{G}})$ be the jth design point on which no replicated observation lies, and assume that $t_1 < t_2 < \cdots < t_s$ and $s = \#K_{\mathscr{G}}$ without loss of generality. Then we have

$$(6.15) \qquad \sum_{i \in K_{\mathscr{G}}} (x_{i^+} - x_{i^-})^2 \le \sum_{j=2}^{s-1} (t_{j+1} - t_{j-1})^2$$

$$\leq 2\sum_{j=2}^{s-1}\{(t_{j+1} - t_j)^2 + (t_j - t_{j-1})^2\} \leq 4\sum_{j=1}^{s-1}(t_{j+1} - t_j)^2$$

$$\leq 4(t_s - t_1)^2 \leq 4(b-a)^2 < +\infty,$$

and

(6.16) $$\sum_{i\in K}(x_{i^+} - x_{i^-})^2(x_i - x_{i^-})^2 \leq 2^{-4}\sum_{i\in K_{\mathcal{G}}}(x_{i^+} - x_{i^-})^4$$

$$\leq \sum_{j=2}^{s-1}(t_{j+1} - t_{j-1})^4 \leq (b-a)^4 < +\infty.$$

Hence, the assertion follows from COROLLARIES 6.2 and 6.3.

Following similar lines as in the general theory in Section 5, we obtain the following theorem:

THEOREM 6.2. Suppose that CONDITION 3* in Section 5 holds. Then, the asymptotic distribution of $\sqrt{\nu_G}(\hat{\Sigma}_G - \Sigma)$ is normal with mean $\underset{\sim}{0}$ and covariances (5.8) if $\sum_{i\in K_{\mathcal{G}}} \underset{\sim}{\xi}_i' \underset{\sim}{\xi}_i = o(\sqrt{n})$ as $n \to +\infty$.

In the proof of THEOREM 6.2 the following lemma is essential.

LEMMA 6.4. Let V_G be the matrix given by REMARK 6.1, and let $\lambda_1(V_G)$ be the largest eigenvalue of V_G. Then

(6.17) $$\frac{1}{2} - \frac{1}{n} \leq \lambda_1(V_G) \leq \frac{17}{4}.$$

Proof. Note that $\lambda_1(V_G) = \sup_{\underset{\sim}{u}'\underset{\sim}{u}=1} \sum_\alpha \sum_\beta v_{\alpha\beta} u_\alpha u_\beta$, where $v_{\alpha\beta}$'s are given in REMARK 6.1. After some straightforward calculations, we can show that $\lambda_1(V_G) \leq 17/4$. The left hand part of (6.17) follows from

$$\frac{1}{2} - \frac{1}{n} \le \frac{1}{2}(n-2)/n \le \sum_{i \in K} c_i^2/n = \mathrm{tr}V_G/n \le \lambda_1(V_G).$$

Proof of THEOREM 6.2. From LEMMA 6.4, we see that CONDITION 4 in Section 5 is automatically satisfied. Hence, the assertion follows from THEOREM 5.2.

Using arguments similar to the ones in deriving COROLLARIES 6.1, 6.2 and 6.3 from THEOREM 6.1, we obtain the following corollaries of THEOREM 6.2:

COROLLARY 6.4. Suppose that CONDITION $3^{\dagger}$ holds, and that the same conditions as in COROLLARY 6.1 hold. Then the asymptotic distribution of $\sqrt{\nu_{G\mathfrak{U}}}(\hat{\Sigma}_{G\mathfrak{U}} - \Sigma)$ is normal with mean $\underset{\sim}{0}$ and covariances (5.8) if $\sum_{i \in K_{\mathscr{G}}} (x_{i^+} - x_{i^-})^2 = o(\sqrt{n})$ as $n \to +\infty$.

COROLLARY 6.5. Suppose that CONDITION $3^{\dagger}$ holds, and that the same conditions as in COROLLARY 6.2 hold. Then the asymptotic distribution of $\sqrt{\nu_{G\mathscr{L}}}(\hat{\Sigma}_{G\mathscr{L}} - \Sigma)$ is normal with mean $\underset{\sim}{0}$ and covariances (5.8) if $\sum_{i \in K_{\mathscr{G}}} (x_{i^+} - x_i)^2 (x_i - x_{i^-})^2 = o(\sqrt{n})$ as $n \to +\infty$.

COROLLARY 6.6. Suppose that CONDITION $3^{\dagger}$ holds, and that there exist two numbers a and b such that $-\infty < a \le x_i \le b < +\infty$ for all $i \in K$. Then, the asymptotic distribution of $\sqrt{\nu_{G\mathfrak{U}}}(\hat{\Sigma}_{G\mathfrak{U}} - \Sigma)$ is normal with mean $\underset{\sim}{0}$ and covariances (5.8) if $\underset{\sim}{\eta}$ is differentiable on [a, b]; so is also that of $\sqrt{\nu_{G\mathscr{L}}}(\hat{\Sigma}_{G\mathscr{L}} - \Sigma)$ if $\underset{\sim}{\eta}$ is twice differentiable on [a, b].

7. Testing goodness of fit of linear models

In this section we propose a criterion for testing goodness of fit of linear models in multivariate regression. Assume that the regression relation can be described as in the model (1.1) and that the errors $\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2 \cdots$ are independently distributed according to $N_p(\underset{\sim}{0}, \Sigma)$.

Suppose that a hypothesized model, say f-Model, is expressed as

$$(7.1) \qquad \P = X_f\theta,$$

where X_f is an $n \times r$ design matrix induced by a function $\underset{\sim}{f} = (f_1, \cdots, f_r)': R^q \to R^r$, that is

$$(7.2) \qquad X_f = \begin{bmatrix} \underset{\sim}{f}(\underset{\sim}{x}_1)' \\ \vdots \\ \underset{\sim}{f}(\underset{\sim}{x}_n)' \end{bmatrix} = [\underset{\sim}{f}^{(1)}, \cdots, \underset{\sim}{f}^{(r)}],$$

where the function $\underset{\sim}{f}$ is known, $\text{rank}(X_f) = r$ and θ is an unknown $r \times p$ coeficient matrix. When there are enough replicated observations in data set, it is possible to test the hypothesis H_f: "Model f is true" by using the Wilks' Λ-statistics (Wilks[24]) derived below.

Let $\hat{\Sigma}_{PE} = Y'V_{PE}Y/(n-g)$ be the PEMS estimator defined by (1.3), where g is the total number of distinct design points in the data. Here we assume that $n - g \geq p$, and let $\hat{\Sigma}_f = Y'(I_n - P_f)Y/(n-r)$, where $P_f = X_f(X_f'X_f)^{-1}X_f'$. From the general theory of multivariate linear model (see, e.g., Anderson[1], Seber[19], Siotani et al.[21]), the likelihood ratio criterion is based on

$$(7.3) \qquad \Lambda = \frac{|(n-g)\hat{\Sigma}_{PE}|}{|(n-g)\hat{\Sigma}_{PE} + \{(n-r)\hat{\Sigma}_f - (n-g)\hat{\Sigma}_{PE}\}|} = \frac{|\hat{\Sigma}_{PE}|}{|\hat{\Sigma}_f|} \cdot \frac{n-g}{n-r}.$$

Under H_f, $(n-g)\hat{\Sigma}_{PE}$ and $(n-r)\hat{\Sigma}_f - (n-g)\hat{\Sigma}_{PE}$ have independent Wishert distributions $W_p(n-g, \Sigma)$ and $W_p(g-r, \Sigma)$, respectively. Then Λ has a Λ-distribution with degrees of freedom p, g-r, and n-g. For the tables of the upper quantile values for the Λ-distribution, see, e.g., Seber[19]. If the ratio $|\hat{\Sigma}_{PE}|/|\hat{\Sigma}_f|$ is very smaller than the expected value under H_f, that is, if $|\hat{\Sigma}_f|$ is much greater than $|\hat{\Sigma}_{PE}|$, we reject H_f and may suspect that there exist some lack of fit in f-Model. It is noteworthy that the test based on the Λ-statistic of (7.3) is equivalent to the well-known classical F-test when $p = 1$ (see, e.g., Seber[18, Section 4.4]).

The Λ-test mentioned above, unfortunately, can not be applied if there are few replicated observations in the data set. This is the situation we now consider. One possible approach to such a situation is to use the Λ-statistics defined by replacing $\hat{\Sigma}_{PE}$ by a nonparametric estimator $\hat{\Sigma}_{\mathcal{N}}$; however, no simple expression of the exact distribution even when H_f is true is available for the resulting statistics. We now consider the asymptotic distribution of $|\hat{\Sigma}_f|/|\hat{\Sigma}_{\mathcal{N}}|$ when n is large. It is seen that after multiplying a suitable normalizing constant, $\log\{|\hat{\Sigma}_f|/|\hat{\Sigma}_{\mathcal{N}}|\}$ and $\mathrm{tr}(\hat{\Sigma}_f\hat{\Sigma}_{\mathcal{N}}^{-1}) - p$ have the common asymptotic distribution. So we study the distribution of the latter statistics.

THEOREM 7.1. Suppose that $\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2, \cdots$ are independently normally distributed with mean $\underset{\sim}{0}$ and covariance matrix Σ. Let

$$\kappa_r = (2p)^{-1}\{\nu_{\mathcal{N}}^{-1} - (n-r)^{-1}\}^{-1}, \tag{7.4}$$

where $\nu_{\mathcal{N}} = (\mathrm{tr}V_{\mathcal{N}})^2/\mathrm{tr}V_{\mathcal{N}}^2$. Then under H_f the asymptotic distribution of

$$T_f = \sqrt{\kappa_r}\{\mathrm{tr}(\hat{\Sigma}_f\hat{\Sigma}_{\mathcal{N}}^{-1}) - p\} \tag{7.5}$$

is N(0, 1) if the following conditions are fulfilled:

(i) $$\nu_{\mathcal{N}} = (\mathrm{tr}V_{\mathcal{N}})^2/\mathrm{tr}V_{\mathcal{N}}^2 \to +\infty, \quad \text{as } n \to +\infty,$$

and $$\limsup_{n \to +\infty} \nu_{\mathcal{N}}/(n-r) < 1.$$

(ii) There exists a positive number G such that

$$\max_{i \in K} n_i \|\underset{\sim}{w}_i\|^2 \le G < +\infty.$$

(iii) $$\lambda_1(V_{\mathcal{N}}) = o(\sqrt{n_K}) \quad \text{as } n \to +\infty.$$

(iv) There exists a positive definite matrix Ω_f such that

$$X_f'X_f/n \to \Omega_f, \qquad \text{as } n \to +\infty.$$

(v) Let $\underset{\sim}{e}_i = X_f'(\underset{\sim}{w}_i - \underset{\sim}{\delta}_i)$, $i \in K$. Then

$$\sum_{i \in K} \underset{\sim}{e}_i'\underset{\sim}{e}_i = o(\sqrt{n_k}), \qquad \text{as } n \to +\infty.$$

Proof. Note that $T_f = \sqrt{\kappa_r}\mathrm{tr}[(\hat{\Sigma}_f - \hat{\Sigma}_{\mathcal{N}})\Sigma^{-1}\{I_p + (\hat{\Sigma}_{\mathcal{N}} - \Sigma)\Sigma^{-1}\}^{-1}]$ and from THEOREM 5.1 $\hat{\Sigma}_{\mathcal{N}} \to \Sigma$ as $n \to +\infty$ in probability. Hence, the asymptotic distribution of T_f is the same as that of $T_f^{\dagger} = \sqrt{\kappa_r}\mathrm{tr}\{\Sigma^{-1}(\hat{\Sigma}_f - \hat{\Sigma}_{\mathcal{N}})\}$. Letting

(7.3) $$W_f = (n-r)^{-1}(I_n - P_f) - (\mathrm{tr}V_{\mathcal{N}})^{-1}V_{\mathcal{N}},$$

we have $T_f^{\dagger} = \sqrt{\kappa_r}\mathrm{tr}(\Sigma^{-1}Y'W_fY)$. When the null hypothesis is true, there exists an $r \times p$ matrix θ such that $\P = E[Y] = X_f\theta$, and hence $T_f^{\dagger}$ can be expressed in terms of $\mathcal{E} = Y - X_f\theta$ and decomposed as follows:

(7.4) $$T_f^{\dagger} = \Delta_f + 2\tau_f + T_f^{*},$$

where $$\Delta_f = -\sqrt{\kappa_r}(\mathrm{tr}V_{\mathcal{N}})^{-1}\mathrm{tr}(\theta\Sigma^{-1}\theta'X_f'V_{\mathcal{N}}X_f),$$

$$\tau_f = -\sqrt{\kappa_f}(\mathrm{tr}V_{\mathcal{N}})^{-1}\mathrm{tr}(\Sigma^{-1}\theta'X_f'V_{\mathcal{N}}\mathcal{E}),$$

and $$T_f^{*} = \sqrt{\kappa_r}\mathrm{tr}(\Sigma^{-1}\mathcal{E}'W_f\mathcal{E}).$$

First we show that $\Delta_f \to 0$ as $n \to +\infty$ and $\tau_f \to 0$ as $n \to +\infty$ in probability. Note that $\mathrm{tr}V_{\mathcal{N}} = \sum_{i\in K}(1+\|\underset{\sim}{w}_i\|^2)^{-1} \geq n_K/(1+G)$ from (ii). Using a similar calculation as in (5.2), we obtain from (v) and LEMMA 3.1 that

$$|\Delta_f| \leq \frac{1+G}{\sqrt{2p}}\sum_j\sum_\ell |\underset{\sim}{\theta}_j'\Sigma^{-1}\underset{\sim}{\theta}_\ell|\left(\frac{n-r}{n-r-\nu_{\mathcal{N}}}\right)^{\frac{1}{2}}\left(\sum_{i\in K}\underset{\sim}{e}_i'\underset{\sim}{e}_i/\sqrt{n_K}\right) \to 0,$$

as $n \to +\infty$, where $\theta = [\underset{\sim}{\theta}_1, \cdots, \underset{\sim}{\theta}_r]'$. We also obtain from (A.5) and (5.2) that under (i), (ii), (iii) and (iv)

$$E[\tau_f^2] = \kappa_r(\mathrm{tr}V_{\mathcal{N}})^{-2}\mathrm{tr}(\Sigma^{-1}\theta' X_f' V_{\mathcal{N}}^2 X_f \theta)$$

$$\leq (2p)^{-1}(1+G)^2\{\lambda_1(V_{\mathcal{N}})/\sqrt{n_K}\}\left(\sum_j \sum_\ell |\underset{\sim}{\theta}_j'\Sigma^{-1}\underset{\sim}{\theta}_\ell|\right)\left(\sum_{i\in K}\underset{\sim}{e}_i'\underset{\sim}{e}_i/\sqrt{n_K}\right) \rightarrow 0,$$

as $n \rightarrow +\infty$. This implies that $\tau_f \rightarrow 0$ as $n \rightarrow +\infty$ in probability.

Finally we show that T_f^* is asymptotically distributed according to N(0,1). Let $\phi_n(t)$ be the characteristic function of T_f^*. Then following similar lines as in (5.9), we can obtain

$$\phi_n(t) = E[\exp\{it\sqrt{\kappa_r}\mathrm{tr}(\Sigma^{-1}\mathcal{E}' W_f \mathcal{E})\}]$$

$$= \prod_\alpha \{1 - 2it\sqrt{\kappa_r}\lambda_\alpha(W_f)\}^{-\frac{p}{2}}.$$

Hence

$$\log \phi_n(t) = -\frac{1}{2}\,p\sum_\alpha \log\{1 - 2it\sqrt{\kappa_r}\lambda_\alpha(W_f)\}.$$

Using a Taylor expansion, we have that for any $t \in (-\infty, +\infty)$

$$\log\{1 - 2it\sqrt{\kappa_r}\lambda_\alpha(W_f)\} = -2i\sqrt{\kappa_r}\lambda_\alpha(W_f)t + 2\{\sqrt{\kappa_r}\lambda_\alpha(W_f)\}^2 t^2 + R_\alpha^{(n)}(t),$$

where

$$R_\alpha^{(n)}(t) = \frac{8}{3}\,i\,\frac{\{\sqrt{\kappa_r}\,\lambda_\alpha(W_f)\}^3}{\{1 - 2i\theta_\alpha(t)t\sqrt{\kappa_r}\lambda_\alpha(W_f)\}^3}\,t^3,$$

for some $\theta_\alpha(t)$ in (0, 1). Note that $\mathrm{tr}W_f = 0$ and

$$\mathrm{tr}W_f^2 = \nu_{\mathcal{N}}^{-1} - (n-r)^{-1} + 2(n-r)^{-1}(\mathrm{tr}V_{\mathcal{N}})^{-1}\mathrm{tr}(P_f V_{\mathcal{N}}).$$

Letting $\Omega_{f,n} = X_f' X_f/n$, we have

$$\mathrm{tr}(P_f V_{\mathcal{N}}) = \mathrm{tr}\{(X_f' X_f)^{-1} X_f' V_{\mathcal{N}} X_f\} = n^{-1} \mathrm{tr}(\Omega_{f,n}^{-1} X_f' V_{\mathcal{N}} X_f)$$

$$= n^{-1} \sum_j \sum_\ell \omega_n^{j\ell} \sum_{i \in K} c_i^4 e_{ij} e_{i\ell},$$

where $\omega_n^{j\ell}$ is the (j,ℓ)-element of the inverse of $\Omega_{f,n}$. Hence, we obtain from (i), (ii), (iv) and (v) that

$$(7.5) \qquad |\kappa_r \mathrm{tr} W_f^2 - (2p)^{-1}| = \frac{\nu_{\mathcal{N}}}{p(n-r-\nu_{\mathcal{N}})\mathrm{tr} V_{\mathcal{N}}} \cdot |\mathrm{tr}(P_f V_{\mathcal{N}})|$$

$$\leq \frac{1+G}{p\sqrt{n_K}} \left(\frac{\nu_{\mathcal{N}}}{n-r-\nu_{\mathcal{N}}}\right) \left(\sum_j \sum_\ell |\omega_n^{j\ell}|\right) \left(\sum_{i \in K} \underset{\sim}{e}_i' \underset{\sim}{e}_i / \sqrt{n_K}\right) = o(1/\sqrt{n_K}),$$

as $n \to +\infty$. Therefore, letting $\delta_n(t) = \frac{p}{2} \sum_\alpha R_\alpha^{(n)}(t)$, we obtain

$$\log \phi_n(t) = ip\sqrt{\kappa_r}(\mathrm{tr} W_f)t - \kappa_r p(\mathrm{tr} W_f^2)t^2 + \delta_n(t)$$

$$= -\{\tfrac{1}{2} + o(1/\sqrt{n_K})\}t^2 + \delta_n(t).$$

Since, under Conditions (ii) and (iii),

$$\sqrt{\kappa_r} \max_\alpha |\lambda_\alpha(W_f)| \leq \sqrt{\kappa_r}\{\lambda_1(V_{\mathcal{N}})/\mathrm{tr} V_{\mathcal{N}} + (n-r)^{-1}\}$$

$$\leq (1+G)\sqrt{\kappa_r/n_K}\left(\lambda_1(V_{\mathcal{N}})/\sqrt{n_K}\right) + \sqrt{\kappa_r}/(n-r)$$

$$= O(1)\left(\lambda_1(V_{\mathcal{N}})/\sqrt{n_K}\right) + O(1/\sqrt{n}) \to 0,$$

as $n \to +\infty$, it follows from (7.5) that

$$|\delta_n(t)| \leq \frac{4}{3} p \, \kappa_r^{\frac{3}{2}} \left(\sum_\alpha |\lambda_\alpha(W_f)|^3\right) t^3$$

$$= \frac{4}{3}\{p\kappa_r(\mathrm{tr}W_f^2)\}\left(\sqrt{\kappa_r}\max_\alpha|\lambda_\alpha(W_f)|\right) \to 0,$$

as $n \to +\infty$. Hence, we have that $\phi_n(t) \to \exp\left(-\frac{1}{2}t^2\right)$ for any fixed t. This completes the proof.

COROLLARY 7.1. Suppose that every component of $\underset{\sim}{f} = (f_1, \cdots, f_r)'$ is differentiable and satisfies

$$\psi_\alpha = \sup_{\underset{\sim}{t}} \left\{\sum_{j=1}^{q}\left(\frac{\partial}{\partial x_j}f_\alpha(\underset{\sim}{x})\Big|_{\underset{\sim}{x}=\underset{\sim}{t}}\right)^2\right\}^{\frac{1}{2}} < +\infty, \quad \alpha = 1, \cdots, r,$$

and that the conditions (i), (iii) and (iv) in THEOREM 7.1 are fulfilled. Let $\hat{\Sigma}_{\mathfrak{U}}$ be a LUW estimator of Σ, and let $\kappa_{\mathfrak{U},r} = (2p)^{-1}\nu_{\mathfrak{U}}(n-r)/(n-r-\nu_{\mathfrak{U}})$ and $\nu_{\mathfrak{U}} = (\mathrm{tr}V_{\mathfrak{U}})^2/\mathrm{tr}V_{\mathfrak{U}}^2$. If $\sum_{i\in K} d_i^2 = o(\sqrt{n_K})$, as $n \to +\infty$, then the asymptotic distribution of $T_{\mathfrak{U},f} = \sqrt{\kappa_{\mathfrak{U},r}}\{\mathrm{tr}(\hat{\Sigma}_f\hat{\Sigma}_{\mathfrak{U}}^{-1}) - p\}$ is N(0, 1) when the hypothesis H_f is true.

Proof. For a LUW estimator $\hat{\Sigma}_{\mathfrak{U}}$, we have that $n_i\|\underset{\sim}{w}_i\|^2 = 1$ for all $i \in K$, and obtain from LEMMA 3.1 that

$$0 \le \sum_{i\in K}\underset{\sim}{e}_i'\underset{\sim}{e}_i/\sqrt{n_K} \le \left(\sum_{\alpha=1}^{r}\psi_\alpha^2\right)\left(\sum_{i\in K}d_i^2/\sqrt{n_K}\right) \to 0,$$

as $n \to +\infty$. Hence, the assertion follows from THEOREM 7.1.

COROLLARY 7.2. Suppose that every component of $\underset{\sim}{f} = (f_1, \cdots, f_r)'$ is twice differentiable and has the Hessian satisfying

$$\gamma_\alpha = \sup_{\underset{\sim}{x}}\sup_{\underset{\sim}{u}'\underset{\sim}{u}=1}|\underset{\sim}{u}'H_{\underset{\sim}{x}}^{(\alpha)}\underset{\sim}{u}| < +\infty, \qquad \alpha = 1, \cdots, r,$$

and that the conditions (i), (ii), (iii) and (iv) in THEOREM 7.1 are fulfilled. Let $\hat{\Sigma}_{\mathscr{L}}$ be a LLW estimator of Σ and let $\kappa_{\mathscr{L},r} = (2p)^{-1}\nu_{\mathscr{L}}(n-r)/(n-r-\nu_{\mathscr{L}})$, where $\nu_{\mathscr{L}} = (\mathrm{tr}V_{\mathscr{L}})^2/\mathrm{tr}V_{\mathscr{L}}^2$. If $\sum_{i\in K} d_i^4 = o(\sqrt{n_K})$, as $n \to +\infty$, then the asymptotic distribution of $T_{\mathscr{L},f} = \sqrt{\kappa_{\mathscr{L},r}}\{\mathrm{tr}(\hat{\Sigma}_f\hat{\Sigma}_{\mathscr{L}}^{-1}) - p\}$ is N(0, 1) when the hypothesis H_f is true.

Proof. From the condition (ii) and LEMMA 3.2 we obtain that

$$0 \le \sum_{i\in K} \underset{\sim}{e}_i'\underset{\sim}{e}_i/\sqrt{n_K} \le \frac{1}{4} G\left(\sum_{\alpha=1}^{r} \gamma_\alpha^2\right)\left(\sum_{i\in K} d_i^4/\sqrt{n_K}\right) \to 0,$$

as $n \to +\infty$. Hence, the assertion follows immediately from THEOREM 7.1.

8. Robust estimators of diagonal elements of Σ

A disadvantage of $\hat{\Sigma}_{\mathscr{N}}$ is that $\hat{\Sigma}_{\mathscr{N}}$ has a lack of robustness because one single outlier may have an arbitrary large effect on the estimator. For diagonal elements of Σ, i.e. variances of the components of $\underset{\sim}{y}$, using the jth components r_{ij} of $\underset{\sim}{r}_i = \hat{\underset{\sim}{y}}_i - \underset{\sim}{y}_i$ $(i \in K)$, and applying the idea due to Rousseeuw[17], we may construct a robust alternative estimator $\hat{\sigma}_{\mathscr{R}}(j,j)$. The derivation of the robust estimator is based on an averaging procedure through taking the median of $c_i^2r_{ij}^2$'s $(i \in K)$ rather than the arithmetic mean of them (see, Hampel[10, p.380]). When errors are normally distributed, the robust alternative may be given by

$$\hat{\sigma}_{\mathscr{R}}(j,j) = 2.198 \ \underset{i\in K}{\mathrm{median}}(c_i^2r_{ij}^2). \tag{8.1}$$

Here $\{1/\Phi^{-1}(3/4)\}^2 \simeq 2.198$ is an asymptotic correction factor, because

$$\underset{i \in K}{\text{median}}(c_i^2 r_{ij}^2) \rightarrow \sigma_{jj} \text{median}(\chi_1^2) = \sigma_{jj}\{\Phi^{-1}(3/4)\}^2,$$

as $n \rightarrow +\infty$, where Φ denotes the standard normal distribution function.

Another robust alternative may be given by an M-estimator which was introduced by Huber[11]. The scale M-estimator for the jth element of Σ is defined in our case as follows. Let ρ be a real function satisfying the following assumptions.

(i) $\rho(0) = 0$;

(ii) $\rho(-u) = \rho(u)$;

(iii) $0 \le u \le v$ implies $\rho(u) \le \rho(v)$;

(iv) ρ is continuous;

(v) $0 < a = \sup \rho(u) < +\infty$;

(vi) if $\rho(u) < a$ and $0 \le u < v$, then $\rho(u) < \rho(v)$.

Then, the M-estimator of σ_{jj}, say $\hat{\sigma}_{\mathcal{M}}(j,j)$, is defined as the value of s which is the solution of

$$n_K^{-1} \sum_{i \in K} \rho(\sqrt{c_i^2} r_{ij}/s) = b,$$

where b may be defined by $E_{\Phi}(\rho(u)) = b$.

The degree of robustness of an estimator in the presence of outliers may be measured by the concept of breakdown-point which was introduced by Hample[9]. Donoho and Huber[6] gave a finite sample version of this concept which will be used here. The finite sample breakdown-point measures the maximum fraction of outliers which a given sample may contain without spoiling the estimator completely.

THEOREM 8.1. Let $\hat{\sigma}_{\mathfrak{R}}(j,j)$ be the estimator given by (8.1). Let U_n be the quantity given by (4.7), and let g be the number of distinct design points in data. Then the breakdown-point of $\hat{\sigma}_{\mathfrak{R}}(j,j)$ is no less than

$$\left[\left[\frac{g}{2}\right]^{+}/(1 + U_n)\right]^{-}/n,$$

where $[t]^{+}$ and $[t]^{-}$ denote the operations of rasing to a unit and of omitting fractions on a real number t, respectively.

Proof. Let $m^{\dagger}$ be the total number of outliers. Then the number of affected elements of $\{c_i^2 r_{ij}^2 \mid i \in K\}$ is at most $(1 + U_n)m^{\dagger}$. From the definition we see easily that $\hat{\sigma}_{\mathfrak{R}}(j,j)$ can not take arbitray large value when $(1 + U_n)m^{\dagger} \le \frac{g}{2} - 1$ or $\frac{g-1}{2}$ according as g is even or odd. Now the assertion follows immediately.

In one-dimensional regression, i.e. the case of q = 2, with no replicated observations in the data, the breakdown-point of $\hat{\sigma}_{\mathfrak{R}}(j,j)$ is $\left[\frac{n^*}{6}\right]^{+}/n$, where n^* is n-2 if n is even, n-1 if n is odd. Hence the asymptotic value is $\frac{1}{6}$.

Appendix. Covariances of some qudaratic forms

Let $Y = [\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_n]'$ be an $n \times p$ random matrix such that $\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_n$ are independently distributed with means $\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_n$, common covariance matrix Σ and common third and fourth moments about their means. The common third and fourth moments are expressed by $\mu_3(j,k,\ell)$

and $\mu_4(j,k,\ell,m)$ for $1 \le j,\ k,\ \ell,\ m \le p$, respectively, as in (4.1) and (4.2).

THEOREM A.1. If $A = [a_{j\ell}]$ and $B = [b_{j\ell}]$ are any $p \times p$ symmetric matrices, $V = [v_{\alpha\beta}]$ is any $n \times n$ symmetric matrix, then

$$\text{(A.1)} \qquad \mathrm{Cov}[\mathrm{tr}(AY'VY),\ \mathrm{tr}(BY'VY)]$$

$$= \underset{\sim}{v}'\underset{\sim}{v}\{\sum_j \sum_k \sum_\ell \sum_m a_{jk}b_{\ell m}\mu_4(j,k,\ell,m) - \mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma) - 2\mathrm{tr}(A\Sigma B\Sigma)\}$$

$$+ \quad 2(\mathrm{tr}V^2)\mathrm{tr}(A\Sigma B\Sigma)$$

$$+ \quad 2\sum_j \sum_k \sum_\ell \sum_m a_{jk}b_{\ell m}\{\mu_3(k,\ell,m)\underset{\sim}{\eta}^{(j)\prime} + \mu_3(m,j,k)\underset{\sim}{\eta}^{(\ell)\prime}\}V\underset{\sim}{v}$$

$$+ \quad 4\mathrm{tr}(A\Sigma B\P' V^2 \P),$$

where $\P = (\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_n)'$ and $\underset{\sim}{v}$ is the column vector of the diagnal elements of V.

Proof. Letting $\mathcal{E} = (\underset{\sim}{\varepsilon}_1, \cdots, \underset{\sim}{\varepsilon}_n)' = Y - \P$, we have

$$\mathrm{tr}(AY'VY) = \mathrm{tr}[A(\P + \mathcal{E})'V(\P + \mathcal{E})]$$

$$= \mathrm{tr}(A\P'V\P) + 2\mathrm{tr}(A\P'V\mathcal{E}) + \mathrm{tr}(A\mathcal{E}'V\mathcal{E}).$$

Note that $E[\mathcal{E}] = O_{n\times p}$ and $E[\underset{\sim}{\varepsilon}^{(j)}\underset{\sim}{\varepsilon}^{(\ell)\prime}] = \sigma_{j\ell}I_n$ for $1 \le j,\ \ell \le p$, $E[\underset{\sim}{\varepsilon}_\alpha \underset{\sim}{\varepsilon}_\beta'] = \delta_{\alpha\beta}\Sigma$ for $1 \le \alpha,\ \beta \le n$, and the third and fourth moments are given by (4.1) and (4.2), respectively. The expectations are calculated in terms of $\mathcal{E}$ and their computations are straightforward.

Here we list some fundamental results in the following:

(A.2) $$E[\mathrm{tr}(AY'VY)] = (\mathrm{tr}V)\mathrm{tr}(A\Sigma) + \mathrm{tr}(A\P'V\P).$$

(A.3) $$E[\mathrm{tr}(V\mathcal{E}'A\mathcal{E})\mathrm{tr}(V\mathcal{E}'B\mathcal{E})] = \underset{\sim}{v}'\underset{\sim}{v}\{\sum_j \sum_k \sum_\ell \sum_m a_{jk}b_{\ell m}\mu_4(j,k,\ell,m)$$

$$- \mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma) - 2\mathrm{tr}(A\Sigma B\Sigma)\}$$

$$+ 2(\mathrm{tr}V^2)\mathrm{tr}(A\Sigma B\Sigma) + (\mathrm{tr}V)^2\{\mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma)\}.$$

(A.4) $$E[\mathrm{tr}(A\P'V\mathcal{E})\mathrm{tr}(B\mathcal{E}'V\mathcal{E})]$$

$$= \sum_j \sum_k \sum_\ell \sum_m a_{jk}b_{\ell m}\mu_3(k,\ell,m)\underset{\sim}{v}'V\underset{\sim}{\eta}^{(j)}.$$

(A.5) $$E[\mathrm{tr}(A\P'V\mathcal{E})\mathrm{tr}(B\P'V\mathcal{E})] = \mathrm{tr}(A\Sigma B\P'V^2\P).$$

COROLLARY A.1. If $\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_n$ are also normally distributed in THEOREM A.1, then $\mu_3(j,k,\ell) = 0$, $\mu_4(j,k,\ell,m) = \sigma_{jk}\sigma_{\ell m} + \sigma_{j\ell}\sigma_{km} + \sigma_{jm}\sigma_{\ell k}$, and

$$\mathrm{Cov}[\mathrm{tr}(AY'VY), \mathrm{tr}(BY'VY)] = 2(\mathrm{tr}V^2)\mathrm{tr}(A\Sigma B\Sigma) + 4\mathrm{tr}(A\Sigma B\P'V^2\P).$$

COROLLARY A.2. Let $p = 1$ in THEOREM A.1. Then, we obtain the following well-known expression of variance of a quadratic form (see,e.g., Atiqullah[2], Seber[18, Chapter 1.4]):

$$\mathrm{Var}[Y'VY] = \underset{\sim}{v}'\underset{\sim}{v}(\mu_4 - 3\sigma^4) + 2(\mathrm{tr}V^2)\sigma^4$$

$$+ 4\mu_3\, \underset{\sim}{\eta}'V\underset{\sim}{v} + 4\sigma^2\underset{\sim}{\eta}'V^2\underset{\sim}{\eta}.$$

Acknowledgements

The author would like to express his deep gratitude to Prof. Y. Fujikoshi, Hiroshima University, his valuable advices and encouragement, and also Prof. M. Munaka, Hiroshima University, for his support and encouragement.

References

[1] T. W. Anderson, Introduction to Multivariate Statistical Analysis, 2nd eds., Wiley, New York, 1984.

[2] M. Atiqullah, The estimation of residual variance in quadratically balanced least squares problems and the robustness of the F-test, Biometrika **49** (1962), 83-91.

[3] M. J. Buckley, G. K. Eagleson and B. W. Silverman, The estimation of residual variance in nonparametric regression, Biometrika **75**, (1988), 189-199.

[4] W. S. Cleveland and S. J. Devlin, Locally weighted regression: An approach to regression analysis by local fitting, J. Am. Stat. Soc. **83**, (1988), 596-622.

[5] C. Daniel and F. S. Wood, Fitting Equations to Data, John Wiley, New York, 1971.

[6] D. L. Donoho and P. J. Huber, Notion of Breakdown Point, in A

Festschrift for Erich L. Lehman, eds. P. Bickel, K. Doksum, and J. L. Hodges, Jr., Belmont, Wadsworth, Calif., 1983.

[7] N. R. Draper and H. Smith, Applied Regression Analysis, 2nd eds., John Wiley, New York, 1981.

[8] T. Gasser, L. Sroka and C. Jenneu-Steimmetz, Residual variance and residual pattern in nonlinear regression, Biometrika **73**, (1986), 623-633.

[9] F. R. Hampel, A general qualitative definition of robustness, Ann. Math. Statist. **42**, (1971), 1887-1896.

[10] F. R. Hampel, Beyond location parameters: robust concepts an methods, Bull. Inter. Statist. Inst. **46**, (1975), 375-382.

[11] P. J. Huber, Robust Statistics, Wiley, New York, 1981.

[12] J. M. Landwehr, D. Pregibon and A. C. Shoemaker, Graphical methods for assessing logistic regression models, J. Am. Stat. Assoc. **79**, (1984), 61-71.

[13] M. Ohtaki, On the application of linear models to local regions in regression, Statistical Theory and Data Analysis, ed. by K. Matsusita, Elsevier Science Publishers B. V. (North Holland), Amsterdam, (1985), 529-545.

[14] M. Ohtaki, A nonparametric estimate of residual variance in one-dimensional regression, Bull. Biometric Soc. Jpn. **8**, (1987), 39-51.

[15] C. R. Rao, Linear Statistical Inference and Its Applications, 2nd eds., John Wiley, New York, 1973.

[16] J. Rice, Bandwidth choice for nonparametric regression, Ann. Statist. **12**, (1984), 1215-1230.

[17] P. J. Rousseeuw, Least median of squares regression, J. Am. Stat. Assoc. **79**, (1984), 871-880.

[18] G. A. F. Seber, Linear Regression Analysis, John Wiley, New York, 1977.

[19] G. A. F. Seber, Multivariate Observations, John Wiley, New York, 1984.

[20] B. W. Silverman, Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). J. R. Statist. Soc. B **47**, (1985), 1-52.

[21] M. Siotani, T. Hayakawa and Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, Inc., Columbus, Ohio, 1985.

[22] C. J. Stone, Consistent nonparametric regression, Ann. Statist. **5**, (1977), 595-645.

[23] S. Weisberg, Applied Linear Regression, 2nd ed., John Wiley, New York, 1985.

[24] S. S. Wilks, Certain generalizations in the analysis of variance, Biometrika **24**, (1932), 471-494.

Department of Biometrics,
Research Institute for Nuclear Medicine
and Biology, Hiroshima University