Stat 709: Mathematical Statistics
Lecture 38

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 38: Asymptotic properties of LSE's and weighted LSE's

Theorem 3.11 (Consistency)

Consider model

$$X = Z\beta + \varepsilon \tag{1}$$

under assumption A3 ($E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon)$ is an unknown matrix).
Consider the LSE $l^\tau\widehat{\beta}$ with $l \in \mathscr{R}(Z)$ for every n.
Suppose that $\sup_n \lambda_+[\mathrm{Var}(\varepsilon)] < \infty$, where $\lambda_+[A]$ is the largest eigenvalue of the matrix A, and that $\lim_{n\to\infty} \lambda_+[(Z^\tau Z)^-] = 0$.
Then $l^\tau\widehat{\beta}$ is consistent in mse for any $l \in \mathscr{R}(Z)$.

Proof

The result follows from the fact that $l^\tau\widehat{\beta}$ is unbiased and

$$\mathrm{Var}(l^\tau\widehat{\beta}) = l^\tau(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) Z(Z^\tau Z)^- l \le \lambda_+[\mathrm{Var}(\varepsilon)] l^\tau (Z^\tau Z)^- l.$$

Lecture 38: Asymptotic properties of LSE's and weighted LSE's

Theorem 3.11 (Consistency)

Consider model

$$X = Z\beta + \varepsilon \qquad (1)$$

under assumption A3 ($E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon)$ is an unknown matrix).
Consider the LSE $l^\tau\widehat{\beta}$ with $l \in \mathscr{R}(Z)$ for every n.
Suppose that $\sup_n \lambda_+[\mathrm{Var}(\varepsilon)] < \infty$, where $\lambda_+[A]$ is the largest eigenvalue of the matrix A, and that $\lim_{n\to\infty} \lambda_+[(Z^\tau Z)^-] = 0$.
Then $l^\tau\widehat{\beta}$ is consistent in mse for any $l \in \mathscr{R}(Z)$.

Proof

The result follows from the fact that $l^\tau\widehat{\beta}$ is unbiased and

$$\mathrm{Var}(l^\tau\widehat{\beta}) = l^\tau(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) Z(Z^\tau Z)^- l \le \lambda_+[\mathrm{Var}(\varepsilon)] l^\tau (Z^\tau Z)^- l.$$

Without the normality assumption on ε, the exact distribution of $l^\tau\widehat{\beta}$ is very hard to obtain.
The asymptotic distribution of $l^\tau\widehat{\beta}$ is derived in the following result.

Theorem 3.12

Consider model (1) with assumption A3.
Suppose that $0 < \inf_n \lambda_-[\mathrm{Var}(\varepsilon)]$, where $\lambda_-[A]$ is the smallest eigenvalue of the matrix A, and that

$$\lim_{n\to\infty} \max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i = 0. \qquad (2)$$

Suppose further that $n = \sum_{j=1}^k m_j$ for some integers k, m_j, $j = 1,...,k$, with m_j's bounded by a fixed integer m, $\varepsilon = (\xi_1,...,\xi_k)$, $\xi_j \in \mathscr{R}^{m_j}$, and ξ_j's are independent.

(i) If $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$, then for any $l \in \mathscr{R}(Z)$,

$$l^\tau(\widehat{\beta}-\beta)\Big/\sqrt{\mathrm{Var}(l^\tau\widehat{\beta})} \to_d N(0,1). \qquad (3)$$

(ii) Result (3) holds for any $l \in \mathscr{R}(Z)$ if, when $m_i = m_j$, $1 \le i < j \le k$, ξ_i and ξ_j have the same distribution.

Without the normality assumption on ε, the exact distribution of $l^\tau\widehat{\beta}$ is very hard to obtain.
The asymptotic distribution of $l^\tau\widehat{\beta}$ is derived in the following result.

Theorem 3.12

Consider model (1) with assumption A3.
Suppose that $0 < \inf_n \lambda_-[\,\mathrm{Var}(\varepsilon)]$, where $\lambda_-[A]$ is the smallest eigenvalue of the matrix A, and that

$$\lim_{n\to\infty} \max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i = 0. \qquad (2)$$

Suppose further that $n = \sum_{j=1}^k m_j$ for some integers k, m_j, $j = 1,...,k$, with m_j's bounded by a fixed integer m, $\varepsilon = (\xi_1,...,\xi_k)$, $\xi_j \in \mathscr{R}^{m_j}$, and ξ_j's are independent.

(i) If $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$, then for any $l \in \mathscr{R}(Z)$,

$$l^\tau(\widehat{\beta}-\beta)\Big/\sqrt{\mathrm{Var}(l^\tau\widehat{\beta})} \to_d N(0,1). \qquad (3)$$

(ii) Result (3) holds for any $l \in \mathscr{R}(Z)$ if, when $m_i = m_j$, $1 \le i < j \le k$, ξ_i and ξ_j have the same distribution.

Proof

For $l \in \mathscr{R}(Z)$,

$$l^\tau (Z^\tau Z)^- Z^\tau Z \beta - l^\tau \beta = 0$$

and

$$l^\tau (\widehat{\beta} - \beta) = l^\tau (Z^\tau Z)^- Z^\tau \varepsilon = \sum_{j=1}^{k} c_{nj}^\tau \xi_j,$$

where c_{nj} is the m_j-vector whose components are $l^\tau (Z^\tau Z)^- Z_i$, $i = k_{j-1}+1, \ldots, k_j$, $k_0 = 0$, and $k_j = \sum_{t=1}^{j} m_t$, $j = 1, \ldots, k$.
Note that

$$\sum_{j=1}^{k} \|c_{nj}\|^2 = l^\tau (Z^\tau Z)^- Z^\tau Z (Z^\tau Z)^- l = l^\tau (Z^\tau Z)^- l. \qquad (4)$$

Also,

$$\begin{aligned} \max_{1 \le j \le k} \|c_{nj}\|^2 &\le m \max_{1 \le i \le n} [l^\tau (Z^\tau Z)^- Z_i]^2 \\ &\le m l^\tau (Z^\tau Z)^- l \max_{1 \le i \le n} Z_i^\tau (Z^\tau Z)^- Z_i, \end{aligned}$$

which, together with (4) and condition (2), implies that

Proof (continued)

$$\lim_{n\to\infty}\left(\max_{1\le j\le k}\|c_{nj}\|^2 \Big/ \sum_{j=1}^{k}\|c_{nj}\|^2\right)=0.$$

The results then follow from Corollary 1.3.

Remarks

- Under the conditions of Theorem 3.12, $\mathrm{Var}(\varepsilon)$ is a diagonal block matrix with $\mathrm{Var}(\xi_j)$ as the jth diagonal block, which includes the case of independent ε_i's as a special case.
- Exercise 80 shows that condition (2) is almost a necessary condition for the consistency of the LSE.

Lemma 3.3

The following are sufficient conditions for (2).

(a) $\lambda_+[(Z^\tau Z)^-]\to 0$ and $Z_n^\tau(Z^\tau Z)^- Z_n\to 0$, as $n\to\infty$.

(b) There is an increasing sequence $\{a_n\}$ such that $a_n\to\infty$, $a_n/a_{n+1}\to 1$, and $Z^\tau Z/a_n$ converges to a positive definite matrix.

Proof (continued)

$$\lim_{n\to\infty}\left(\max_{1\le j\le k}\|c_{nj}\|^2 \Big/ \sum_{j=1}^{k}\|c_{nj}\|^2\right) = 0.$$

The results then follow from Corollary 1.3.

Remarks

- Under the conditions of Theorem 3.12, $\mathrm{Var}(\varepsilon)$ is a diagonal block matrix with $\mathrm{Var}(\xi_j)$ as the jth diagonal block, which includes the case of independent ε_i's as a special case.
- Exercise 80 shows that condition (2) is almost a necessary condition for the consistency of the LSE.

Lemma 3.3

The following are sufficient conditions for (2).

(a) $\lambda_+[(Z^\tau Z)^-] \to 0$ and $Z_n^\tau (Z^\tau Z)^- Z_n \to 0$, as $n\to\infty$.

(b) There is an increasing sequence $\{a_n\}$ such that $a_n\to\infty$, $a_n/a_{n+1}\to 1$, and $Z^\tau Z/a_n$ converges to a positive definite matrix.

Proof (continued)

$$\lim_{n\to\infty}\left(\max_{1\le j\le k}\|c_{nj}\|^2\Big/\sum_{j=1}^{k}\|c_{nj}\|^2\right)=0.$$

The results then follow from Corollary 1.3.

Remarks

- Under the conditions of Theorem 3.12, $\mathrm{Var}(\varepsilon)$ is a diagonal block matrix with $\mathrm{Var}(\xi_j)$ as the jth diagonal block, which includes the case of independent ε_i's as a special case.
- Exercise 80 shows that condition (2) is almost a necessary condition for the consistency of the LSE.

Lemma 3.3

The following are sufficient conditions for (2).

(a) $\lambda_+[(Z^\tau Z)^-]\to 0$ and $Z_n^\tau(Z^\tau Z)^- Z_n\to 0$, as $n\to\infty$.

(b) There is an increasing sequence $\{a_n\}$ such that $a_n\to\infty$, $a_n/a_{n+1}\to 1$, and $Z^\tau Z/a_n$ converges to a positive definite matrix.

Proof of (a)

Since $Z^\tau Z$ depends on n, we denote $(Z^\tau Z)^-$ by A_n.
Let i_n be the integer such that $h_{i_n} = \max_{1\le i\le n} h_i$.
If $\lim_n i_n = \infty$, then

$$\lim_n h_{i_n} = \lim_n Z_{i_n}^\tau A_n Z_{i_n} \le \lim_n Z_{i_n}^\tau A_{i_n} Z_{i_n} = 0,$$

where the inequality follows from $i_n \le n$ and, thus, $A_{i_n} - A_n$ is nonnegative definite.

If $i_n \le c$ for all n, then

$$\lim_n h_{i_n} = \lim_n Z_{i_n}^\tau A_n Z_{i_n} \le \lim_n \lambda_n \max_{1\le i\le c} \|Z_i\|^2 = 0.$$

Therefore, for any subsequence $\{j_n\} \subset \{i_n\}$ with $\lim_n j_n = a \in (0,\infty]$, $\lim_n h_{j_n} = 0$.
This shows that $\lim_n h_{i_n} = 0$.

Proof of (a)

Since $Z^\tau Z$ depends on n, we denote $(Z^\tau Z)^-$ by A_n.
Let i_n be the integer such that $h_{i_n} = \max_{1\leq i\leq n} h_i$.
If $\lim_n i_n = \infty$, then

$$\lim_n h_{i_n} = \lim_n Z_{i_n}^\tau A_n Z_{i_n} \leq \lim_n Z_{i_n}^\tau A_{i_n} Z_{i_n} = 0,$$

where the inequality follows from $i_n \leq n$ and, thus, $A_{i_n} - A_n$ is nonnegative definite.

If $i_n \leq c$ for all n, then

$$\lim_n h_{i_n} = \lim_n Z_{i_n}^\tau A_n Z_{i_n} \leq \lim_n \lambda_n \max_{1\leq i\leq c} \|Z_i\|^2 = 0.$$

Therefore, for any subsequence $\{j_n\} \subset \{i_n\}$ with $\lim_n j_n = a \in (0,\infty]$, $\lim_n h_{j_n} = 0$.
This shows that $\lim_n h_{i_n} = 0$.

Example: simple linear model

In Example 3.12,

$$X_i = \beta_0 + \beta_1 t_i + \varepsilon_i, \quad i = 1, ..., n.$$

If $n^{-1}\sum_{i=1}^n t_i^2 \to c$ and $n^{-1}\sum_{i=1}^n t_i \to d$ where c is positive and $c > d^2$, then condition (b) in Lemma 3.3 is satisfied with $a_n = n$ and, therefore, Theorem 3.12 applies.

Example: one-way ANOVA

In the one-way ANOVA model (Example 3.13),

$$X_i = \mu_j + \varepsilon_i, \qquad i = k_{j-1} + 1, ..., k_j,\ j = 1, ..., m,$$

where $k_0 = 0$, $k_j = \sum_{l=1}^j n_l$, $j = 1, ..., m$, and $(\mu_1, ..., \mu_m) = \beta$,

$$\max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i = \lambda_+[(Z^\tau Z)^-] = \max_{1\le j\le m} n_j^{-1}.$$

Conditions related to Z in Theorem 3.12 are satisfied iff $\min_j n_j \to \infty$. Some similar conclusions can be drawn in the two-way ANOVA model (Example 3.14).

The weighted LSE

In the linear model

$$X = Z\beta + \varepsilon,$$

the unbiased LSE of $l^\tau\beta$ may be improved by a slightly biased estimator when $V = \mathrm{Var}(\varepsilon)$ is not $\sigma^2 I_n$ and the LSE is not BLUE.

Assume that Z is of full rank so that every $l^\tau\beta$ is estimable.

If V is known, then the BLUE of $l^\tau\beta$ is $l^\tau\check{\beta}$, where

$$\check{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X \qquad (5)$$

(see the discussion after the statement of assumption A3 in §3.3.1).

If V is unknown and $\widehat{V}$ is an estimator of V, then an application of the substitution principle leads to a *weighted least squares estimator*

$$\widehat{\beta}_w = (Z^\tau \widehat{V}^{-1} Z)^{-1} Z^\tau \widehat{V}^{-1} X. \qquad (6)$$

The weighted LSE is not linear in X and not necessarily unbiased for β.

If the distribution of ε is symmetric about 0 and $\widehat{V}$ remains unchanged when ε changes to $-\varepsilon$, then the distribution of $\widehat{\beta}_w - \beta$ is symmetric about 0 and, if $E\widehat{\beta}_w$ is well defined, $\widehat{\beta}_w$ is unbiased for β.

The weighted LSE

In the linear model

$$X = Z\beta + \varepsilon,$$

the unbiased LSE of $l^\tau\beta$ may be improved by a slightly biased estimator when $V = \mathrm{Var}(\varepsilon)$ is not $\sigma^2 I_n$ and the LSE is not BLUE.

Assume that Z is of full rank so that every $l^\tau\beta$ is estimable.
If V is known, then the BLUE of $l^\tau\beta$ is $l^\tau\check{\beta}$, where

$$\check{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X \qquad (5)$$

(see the discussion after the statement of assumption A3 in §3.3.1).

If V is unknown and $\widehat{V}$ is an estimator of V, then an application of the substitution principle leads to a *weighted least squares estimator*

$$\widehat{\beta}_w = (Z^\tau \widehat{V}^{-1} Z)^{-1} Z^\tau \widehat{V}^{-1} X. \qquad (6)$$

The weighted LSE is not linear in X and not necessarily unbiased for β.
If the distribution of ε is symmetric about 0 and $\widehat{V}$ remains unchanged when ε changes to $-\varepsilon$, then the distribution of $\widehat{\beta}_w - \beta$ is symmetric about 0 and, if $E\widehat{\beta}_w$ is well defined, $\widehat{\beta}_w$ is unbiased for β.

The weighted LSE

In the linear model

$$X = Z\beta + \varepsilon,$$

the unbiased LSE of $l^\tau\beta$ may be improved by a slightly biased estimator when $V = \mathrm{Var}(\varepsilon)$ is not $\sigma^2 I_n$ and the LSE is not BLUE.

Assume that Z is of full rank so that every $l^\tau\beta$ is estimable.
If V is known, then the BLUE of $l^\tau\beta$ is $l^\tau\check{\beta}$, where

$$\check{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X \tag{5}$$

(see the discussion after the statement of assumption A3 in §3.3.1).

If V is unknown and $\widehat{V}$ is an estimator of V, then an application of the substitution principle leads to a *weighted least squares estimator*

$$\widehat{\beta}_w = (Z^\tau \widehat{V}^{-1} Z)^{-1} Z^\tau \widehat{V}^{-1} X. \tag{6}$$

The weighted LSE is not linear in X and not necessarily unbiased for β.
If the distribution of ε is symmetric about 0 and $\widehat{V}$ remains unchanged when ε changes to $-\varepsilon$, then the distribution of $\widehat{\beta}_w - \beta$ is symmetric about 0 and, if $E\widehat{\beta}_w$ is well defined, $\widehat{\beta}_w$ is unbiased for β.

If the weighted LSE $l^\tau \widehat{\beta}_w$ is unbiased, then the LSE $l^\tau \widehat{\beta}$ may not be a BLUE, since $\mathrm{Var}(l^\tau \widehat{\beta}_w)$ may be smaller than $\mathrm{Var}(l^\tau \widehat{\beta})$.

Asymptotic properties of the weighted LSE depend on the asymptotic behavior of $\widehat{V}$.

We say that $\widehat{V}$ is consistent for V iff

$$\|\widehat{V}^{-1}V - I_n\|_{\max} \to_p 0, \tag{7}$$

where $\|A\|_{\max} = \max_{i,j} |a_{ij}|$ for a matrix A whose (i,j)th element is a_{ij}.

Theorem 3.17

Consider model (1) with a full rank Z. Let $\breve{\beta}$ and $\widehat{\beta}_w$ be defined by (5) and (6), respectively, with a $\widehat{V}$ consistent in the sense of (7).

Under the conditions in Theorem 3.12,

$$l^\tau(\widehat{\beta}_w - \beta)/a_n \to_d N(0,1),$$

where $l \in \mathscr{R}^p$, $l \neq 0$, and

$$a_n^2 = \mathrm{Var}(l^\tau \breve{\beta}) = l^\tau (Z^\tau V^{-1} Z)^{-1} l.$$

If the weighted LSE $l^\tau \widehat{\beta}_w$ is unbiased, then the LSE $l^\tau \widehat{\beta}$ may not be a BLUE, since $\mathrm{Var}(l^\tau \widehat{\beta}_w)$ may be smaller than $\mathrm{Var}(l^\tau \widehat{\beta})$.

Asymptotic properties of the weighted LSE depend on the asymptotic behavior of $\widehat{V}$.
We say that $\widehat{V}$ is consistent for V iff

$$\|\widehat{V}^{-1}V - I_n\|_{\max} \to_p 0, \tag{7}$$

where $\|A\|_{\max} = \max_{i,j} |a_{ij}|$ for a matrix A whose (i,j)th element is a_{ij}.

Theorem 3.17

Consider model (1) with a full rank Z. Let $\breve{\beta}$ and $\widehat{\beta}_w$ be defined by (5) and (6), respectively, with a $\widehat{V}$ consistent in the sense of (7).
Under the conditions in Theorem 3.12,

$$l^\tau(\widehat{\beta}_w - \beta)/a_n \to_d N(0,1),$$

where $l \in \mathscr{R}^p$, $l \neq 0$, and

$$a_n^2 = \mathrm{Var}(l^\tau \breve{\beta}) = l^\tau (Z^\tau V^{-1} Z)^{-1} l.$$

If the weighted LSE $l^\tau \widehat{\beta}_w$ is unbiased, then the LSE $l^\tau \widehat{\beta}$ may not be a BLUE, since $\mathrm{Var}(l^\tau \widehat{\beta}_w)$ may be smaller than $\mathrm{Var}(l^\tau \widehat{\beta})$.

Asymptotic properties of the weighted LSE depend on the asymptotic behavior of $\widehat{V}$.
We say that $\widehat{V}$ is consistent for V iff

$$\|\widehat{V}^{-1}V - I_n\|_{\max} \to_p 0, \tag{7}$$

where $\|A\|_{\max} = \max_{i,j} |a_{ij}|$ for a matrix A whose (i,j)th element is a_{ij}.

Theorem 3.17

Consider model (1) with a full rank Z. Let $\breve{\beta}$ and $\widehat{\beta}_w$ be defined by (5) and (6), respectively, with a $\widehat{V}$ consistent in the sense of (7).
Under the conditions in Theorem 3.12,

$$l^\tau(\widehat{\beta}_w - \beta)/a_n \to_d N(0,1),$$

where $l \in \mathscr{R}^p$, $l \neq 0$, and

$$a_n^2 = \mathrm{Var}(l^\tau \breve{\beta}) = l^\tau (Z^\tau V^{-1} Z)^{-1} l.$$

Proof

Using the same argument as in the proof of Theorem 3.12, we obtain that

$$l^\tau(\breve{\beta} - \beta)/a_n \to_d N(0,1).$$

By Slutsky's theorem, the result follows from

$$l^\tau \widehat{\beta}_w - l^\tau \breve{\beta} = o_p(a_n).$$

Define

$$\xi_n = l^\tau (Z^\tau \widehat{V}^{-1} Z)^{-1} Z^\tau (\widehat{V}^{-1} - V^{-1})\varepsilon$$

and

$$\zeta_n = l^\tau [(Z^\tau \widehat{V}^{-1} Z)^{-1} - (Z^\tau V^{-1} Z)^{-1}] Z^\tau V^{-1} \varepsilon.$$

Then

$$l^\tau \widehat{\beta}_w - l^\tau \breve{\beta} = \xi_n + \zeta_n.$$

The result follows from $\xi_n = o_p(a_n)$ and $\zeta_n = o_p(a_n)$ (details are in the textbook).

Remarks

- Theorem 3.17 shows that as long as $\widehat{V}$ is consistent in the sense of (7), the weighted LSE $\widehat{\beta}_w$ is asymptotically as efficient as $\breve{\beta}$, which is the BLUE if V is known.
- By Theorems 3.12 and 3.17, the asymptotic relative efficiency of the LSE $l^\tau \widehat{\beta}$ w.r.t. the weighted LSE $l^\tau \widehat{\beta}_w$ is

$$\frac{l^\tau (Z^\tau V^{-1} Z)^{-1} l}{l^\tau (Z^\tau Z)^{-1} Z^\tau V Z (Z^\tau Z)^{-1} l},$$

  which is always less than 1 and equals 1 if $l^\tau \widehat{\beta}$ is a BLUE (in which case $\widehat{\beta} = \breve{\beta}$).
- Finding a consistent $\widehat{V}$ is possible when V has a certain type of structure.

Example 3.29

Consider model (1).
Suppose that $V = \mathrm{Var}(\varepsilon)$ is a block diagonal matrix with the ith diagonal block

$$\sigma^2 I_{m_i} + U_i \Sigma U_i^\tau, \qquad i = 1, ..., k, \tag{8}$$

where m_i's are integers bounded by a fixed integer m, $\sigma^2 > 0$ is an unknown parameter, Σ is a $q \times q$ unknown nonnegative definite matrix, U_i is an $m_i \times q$ full rank matrix whose columns are in $\mathscr{R}(W_i)$, $q < \inf_i m_i$, and W_i is the $p \times m_i$ matrix such that $Z^\tau = (\, W_1 \; W_2 \; ... \; W_k \,)$.

Under (8), a consistent $\widehat{V}$ can be obtained if we can obtain consistent estimators of σ^2 and Σ.
Let $X = (Y_1, ..., Y_k)$, where Y_i is an m_i-vector, and let R_i be the matrix whose columns are linearly independent rows of W_i.
Then

$$\widehat{\sigma}^2 = \frac{1}{n - kq} \sum_{i=1}^{k} Y_i^\tau [I_{m_i} - R_i (R_i^\tau R_i)^{-1} R_i^\tau] Y_i$$

is an unbiased estimator of σ^2.

Example 3.29

Consider model (1).
Suppose that $V = \mathrm{Var}(\varepsilon)$ is a block diagonal matrix with the ith diagonal block

$$\sigma^2 I_{m_i} + U_i \Sigma U_i^\tau, \qquad i = 1, ..., k, \tag{8}$$

where m_i's are integers bounded by a fixed integer m, $\sigma^2 > 0$ is an unknown parameter, Σ is a $q \times q$ unknown nonnegative definite matrix, U_i is an $m_i \times q$ full rank matrix whose columns are in $\mathscr{R}(W_i)$, $q < \inf_i m_i$, and W_i is the $p \times m_i$ matrix such that $Z^\tau = (\ W_1\ W_2\ ...\ W_k\ )$.

Under (8), a consistent $\widehat{V}$ can be obtained if we can obtain consistent estimators of σ^2 and Σ.
Let $X = (Y_1, ..., Y_k)$, where Y_i is an m_i-vector, and let R_i be the matrix whose columns are linearly independent rows of W_i.
Then

$$\widehat{\sigma}^2 = \frac{1}{n - kq} \sum_{i=1}^{k} Y_i^\tau [I_{m_i} - R_i (R_i^\tau R_i)^{-1} R_i^\tau] Y_i$$

is an unbiased estimator of σ^2.

Example 3.29 (continued)

Assume that Y_i's are independent and that $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$ for some $\delta > 0$.

Then $\widehat{\sigma}^2$ is consistent for σ^2 (exercise).

Let $r_i = Y_i - W_i^\tau \widehat{\beta}$ and

$$\widehat{\Sigma} = \frac{1}{k} \sum_{i=1}^{k} \left[(U_i^\tau U_i)^{-1} U_i^\tau r_i r_i^\tau U_i (U_i^\tau U_i)^{-1} - \widehat{\sigma}^2 (U_i^\tau U_i)^{-1} \right].$$

It can be shown (exercise) that $\widehat{\Sigma}$ is consistent for Σ in the sense that $\|\widehat{\Sigma} - \Sigma\|_{\max} \to_p 0$ or, equivalently, $\|\widehat{\Sigma} - \Sigma\| \to_p 0$ (see Exercise 116).