Lecture 32: Asymptotic properties of LSE’s and weighted LSE’s

Theorem 3.11 (Consistency)

Consider model

\[ X = Z\beta + \varepsilon \]  

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

Proof

The result follows from the fact that \(l^\tau\hat{\beta}\) is unbiased and

\[
\text{Var}(l^\tau\hat{\beta}) = l^\tau (Z^\tau Z)^{-} Z^\tau \text{Var}(\varepsilon) Z (Z^\tau Z)^{-} l \leq \lambda_+ [\text{Var}(\varepsilon)] l^\tau (Z^\tau Z)^{-} l.
\]
Without the normality assumption on $\varepsilon$, the exact distribution of $l^\tau \hat{\beta}$ is very hard to obtain.

The asymptotic distribution of $l^\tau \hat{\beta}$ is derived in the following result.

**Theorem 3.12**

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

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

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

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

$$l^\tau (\hat{\beta} - \beta) / \sqrt{\text{Var}(l^\tau \hat{\beta})} \to_d N(0, 1). \quad (3)$$

(ii) Result (3) holds for any $l \in \mathbb{R}(Z)$ if, when $m_i = m_j$, $1 \leq i < j \leq k$, $\xi_i$ and $\xi_j$ have the same distribution.
Proof

For \( l \in \mathcal{R}(Z) \),

\[ l^\tau (Z^\tau Z)^{-1} Z^\tau Z \beta - l^\tau \beta = 0 \]

and

\[ l^\tau (\hat{\beta} - \beta) = l^\tau (Z^\tau Z)^{-1} Z^\tau \epsilon = \sum_{j=1}^{k} c_{nj}^\tau \xi_j, \]

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

Note that

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

Also,

\[ \max_{1 \leq j \leq k} \| c_{nj} \|^2 \leq m \max_{1 \leq i \leq n} [l^\tau (Z^\tau Z)^{-1} Z_i]^2 \]

\[ \leq ml^\tau (Z^\tau Z)^{-1} l \max_{1 \leq i \leq n} Z_i^\tau (Z^\tau Z)^{-1} Z_i, \]

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

\[
\lim_{n \to \infty} \left( \frac{\max_{1 \leq j \leq k} \|c_{nj}\|^2}{\sum_{j=1}^{k} \|c_{nj}\|^2} \right) = 0.
\]

The results then follow from Corollary 1.3.

Remarks

- Under the conditions of Theorem 3.12, \( \text{Var}(\epsilon) \) is a diagonal block matrix with \( \text{Var}(\xi_j) \) as the \( j \)th diagonal block, which includes the case of independent \( \epsilon_i \)'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) \( \lambda_+[(Z^\top Z)^{-}] \to 0 \) and \( Z_n^\top (Z^\top Z)^{-} Z_n \to 0 \), 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 \( Z^\top Z/a_n \) 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 \( h_{i_n} = \max_{1 \leq i \leq n} h_i \).
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,

\[ X_i = \beta_0 + \beta_1 t_i + \varepsilon_i, \quad i = 1, \ldots, n. \]

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),

\[ X_i = \mu_j + \varepsilon_i, \quad i = k_{j-1} + 1, \ldots, k_j, \quad j = 1, \ldots, m, \]

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

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

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

\[ X = Z\beta + \epsilon, \]

the unbiased LSE of \( l^\tau \beta \) may be improved by a slightly biased estimator when \( V = \text{Var}(\epsilon) \) 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 \( l^\tau \hat{\beta} \), where

\[ \hat{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X \]  

(5)

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

If \( V \) is unknown and \( \hat{V} \) is an estimator of \( V \), then an application of the substitution principle leads to a \textit{weighted least squares estimator}

\[ \hat{\beta}_w = (Z^\tau \hat{V}^{-1} Z)^{-1} Z^\tau \hat{V}^{-1} X. \]  

(6)

The weighted LSE is not linear in \( X \) and not necessarily unbiased for \( \beta \). If the distribution of \( \epsilon \) is symmetric about 0 and \( \hat{V} \) remains unchanged when \( \epsilon \) changes to \(-\epsilon\), then the distribution of \( \hat{\beta}_w - \beta \) is symmetric about 0 and, if \( E\hat{\beta}_w \) is well defined, \( \hat{\beta}_w \) is unbiased for \( \beta \).
If the weighted LSE $l^\tau \hat{\beta}_w$ is unbiased, then the LSE $l^\tau \hat{\beta}$ may not be a BLUE, since \( \text{Var}(l^\tau \hat{\beta}_w) \) may be smaller than \( \text{Var}(l^\tau \hat{\beta}) \).

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

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

$$ \| \hat{V}^{-1} V - I_n \|_{\text{max}} \to_p 0, \quad (7) $$

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

**Theorem 3.17**

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

Under the conditions in Theorem 3.12,

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

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

$$ a_n^2 = \text{Var}(l^\tau \check{\beta}) = l^\tau (Z^\tau V^{-1} Z)^{-1} l. $$
Proof

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

\[ l^\tau (\tilde{\beta} - \beta)/a_n \rightarrow_d N(0, 1). \]

By Slutsky’s theorem, the result follows from

\[ l^\tau \hat{\beta}_w - l^\tau \tilde{\beta} = o_p(a_n). \]

Define

\[ \xi_n = l^\tau (Z^\tau \hat{V}^{-1} Z)^{-1} Z^\tau (\hat{V}^{-1} - V^{-1}) \epsilon \]

and

\[ \zeta_n = l^\tau [(Z^\tau \hat{V}^{-1} Z)^{-1} - (Z^\tau V^{-1} Z)^{-1}] Z^\tau V^{-1} \epsilon. \]

Then

\[ l^\tau \hat{\beta}_w - l^\tau \tilde{\beta} = \xi_n + \zeta_n. \]

The result follows from \( \xi_n = o_p(a_n) \) and \( \zeta_n = o_p(a_n) \) (details are in the textbook).
Theorem 3.17 shows that as long as \( \hat{V} \) is consistent in the sense of (7), the weighted LSE \( \hat{\beta}_w \) is asymptotically as efficient as \( \bar{\beta} \), which is the BLUE if \( V \) is known.

By Theorems 3.12 and 3.17, the asymptotic relative efficiency of the LSE \( \ell^\tau \hat{\beta} \) w.r.t. the weighted LSE \( \ell^\tau \hat{\beta}_w \) is

\[
\frac{\ell^\tau (Z^\tau V^{-1} Z)^{-1} \ell}{\ell^\tau (Z^\tau Z)^{-1} Z^\tau V Z (Z^\tau Z)^{-1} \ell},
\]

which is always less than 1 and equals 1 if \( \ell^\tau \hat{\beta} \) is a BLUE (\( \hat{\beta} = \bar{\beta} \)).

Finding a consistent \( \hat{V} \) is possible when \( V \) has a certain type of structure.

**Example 3.29**

Consider model (1).

Suppose that \( V = \text{Var}(\varepsilon) \) is a block diagonal matrix with the \( i \)th diagonal block

\[
\sigma^2 l_{m_i} + U_i \Sigma U_i^\tau, \quad i = 1, \ldots, k,
\]

where \( m_i \)'s are integers bounded by a fixed integer \( m \), \( \sigma^2 > 0 \) is an unknown parameter, \( \Sigma \) is a \( q \times q \) unknown nonnegative definite matrix,
$U_i$ is an $m_i \times q$ full rank matrix whose columns are in $\mathcal{R}(W_i)$, $q < \inf_i m_i$, and $W_i$ is the $p \times m_i$ matrix such that $Z^\tau = (W_1 \; W_2 \; \ldots \; W_k)$. Under (8), a consistent $\hat{V}$ can be obtained if we can obtain consistent estimators of $\sigma^2$ and $\Sigma$.

Let $X = (Y_1, \ldots, Y_k)$, where $Y_i$ is an $m_i$-vector, and let $R_i$ be the matrix whose columns are linearly independent rows of $W_i$.

If $Y_i$’s are independent and $\sup_i E|\varepsilon_i|^2 + \delta < \infty$ for some $\delta > 0$, then

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

is an unbiased and consistent estimator of $\sigma^2$.

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

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

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