ST5215: Advanced Statistical Theory (I)

Chen Zehua

Department of Statistics & Applied Probability

Thursday, November 9, 2011
Theorem 3.11 (Consistency)

Consider model
\[ X = Z\beta + \epsilon \]  
(1)

under assumption A3 (\( E(\epsilon) = 0 \) and \( \text{Var}(\epsilon) \) is an unknown matrix). Consider the LSE \( l^T \hat{\beta} \) with \( l \in \mathcal{R}(Z) \) for every \( n \).

Suppose that \( \sup_n \lambda_+(\text{Var}(\epsilon)) < \infty \), where \( \lambda_+[A] \) is the largest eigenvalue of the matrix \( A \), and that \( \lim_{n \to \infty} \lambda_+[(Z^T Z)^{-}] = 0 \). Then \( l^T \hat{\beta} \) is consistent in mse for any \( l \in \mathcal{R}(Z) \).

Proof

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

\[ \text{Var}(l^T \hat{\beta}) = l^T (Z^T Z)^{-} Z^T \text{Var}(\epsilon) Z (Z^T Z)^{-} l \leq \lambda_+[\text{Var}(\epsilon)] l^T (Z^T Z)^{-} l. \]
Theorem 3.12
Consider model (1) with assumption A3.
Suppose that \( 0 < \inf_n \lambda_- [\text{Var}(\epsilon)] \), where \( \lambda_- [A] \) is the smallest eigenvalue of the matrix \( A \), and that
\[
\lim_{n \to \infty} \max_{1 \leq i \leq n} Z_i^\top (Z^\top Z)^{-1} Z_i = 0. \tag{2}
\]
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 \( m \), \( \epsilon = (\xi_1, ..., \xi_k) \), \( \xi_j \in \mathcal{R}^{m_j} \), and \( \xi_j \)'s are independent.

(i) If \( \sup_i E|\epsilon_i|^{2+\delta} < \infty \), then for any \( l \in \mathcal{R}(Z) \),
\[
\frac{l^\top (\hat{\beta} - \beta)}{\sqrt{\text{Var}(l^\top \hat{\beta})}} \to_d N(0, 1). \tag{3}
\]
(ii) Result (3) holds for any \( l \in \mathcal{R}(Z) \) if, when \( m_i = m_j \), \( 1 \leq i < j \leq k \), \( \xi_i \) and \( \xi_j \) have the same distribution.
Proof
For \( l \in \mathcal{R}(Z) \),
\[
\begin{align*}
I^\top (Z^\top Z)^{-1} Z^\top Z \beta - I^\top \beta &= 0
\end{align*}
\]
and
\[
\begin{align*}
I^\top (\hat{\beta} - \beta) &= I^\top (Z^\top Z)^{-1} Z^\top \epsilon = \sum_{j=1}^k c^T_{nj} \xi_j,
\end{align*}
\]
where \( c_{nj} \) is the \( m_j \)-vector whose components are \( I^\top (Z^\top Z)^{-1} Z_i, \)
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
\[
\begin{align*}
\sum_{j=1}^k \|c_{nj}\|^2 &= I^\top (Z^\top Z)^{-1} Z^\top Z(Z^\top Z)^{-1} I = I^\top (Z^\top Z)^{-1} I. \tag{4}
\end{align*}
\]
Also,
\[
\begin{align*}
\max_{1 \leq j \leq k} \|c_{nj}\|^2 &\leq m \max_{1 \leq i \leq n} [I^\top (Z^\top Z)^{-1} Z_i]^2 \\
&\leq ml^\top (Z^\top Z)^{-1} I \max_{1 \leq i \leq n} Z_i^\top (Z^\top Z)^{-1} Z_i,
\end{align*}
\]
Proof (continued)

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

The results then follow from Corollary 1.3.

- 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 $\lim_{n \to \infty} i_n = \infty$, then

$$
\lim_{n \to \infty} h_{i_n} = \lim_{n \to \infty} Z_{i_n}^\tau A_n Z_{i_n} \leq \lim_{n \to \infty} Z_{i_n}^\tau A_{i_n} Z_{i_n} = 0,
$$

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

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

Therefore, for any subsequence $\{j_n\} \subset \{i_n\}$ with $\lim_{n \to \infty} j_n = a \in (0, \infty]$, $\lim_{n \to \infty} h_{j_n} = 0$.

This shows that $\lim_{n \to \infty} h_{i_n} = 0$. 
Example: simple linear model

In Example 3.12,

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

If \( n^{-1} \sum_{i=1}^{n} t_i^2 \rightarrow c \) and \( n^{-1} \sum_{i=1}^{n} t_i \rightarrow d \) 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),

\[ X_i = \mu_j + \epsilon_i, \quad i = k_j - 1 + 1, \ldots, k_j, \quad j = 1, \ldots, m, \]

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

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

Conditions related to \( Z \) in Theorem 3.12 are satisfied iff

\[ \min_j n_j \rightarrow \infty. \]
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 $l^\tau \tilde{\beta}$, where

$$\tilde{\beta} = (Z^\tau V^{-1} Z)^{-1} Z^\tau V^{-1} X$$  \hspace{1cm} (5)

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

$$\hat{\beta}_w = (Z^\tau \hat{V}^{-1} Z)^{-1} Z^\tau \hat{V}^{-1} X.$$  \hspace{1cm} (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 $l^T \hat{\beta}_w$ is unbiased, then the LSE $l^T \hat{\beta}$ may not be a BLUE, since $\text{Var}(l^T \hat{\beta}_w)$ may be smaller than $\text{Var}(l^T \hat{\beta})$.

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

$$\| \hat{V}^{-1} V - I_n \|_{\text{max}} \rightarrow_p 0,$$

(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 $\tilde{\beta}$ and $\hat{\beta}_w$ be defined by (5) and (6), respectively, with $\hat{V}$ consistent in the sense of (7). Under the conditions in Theorem 3.12,

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

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

$$a_n^2 = \text{Var}(l^T \tilde{\beta}) = l^T (Z^T V^{-1} Z)^{-1} l.$$
Proof
Using the same argument as in the proof of Theorem 3.12, we obtain that
\[ l^\tau (\hat{\beta} - \beta) / a_n \rightarrow_d N(0, 1). \]
By Slutsky’s theorem, the result follows from
\[ l^\tau \hat{\beta}_w - l^\tau \hat{\beta} = o_p(a_n). \]
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
\[ l^\tau \hat{\beta}_w - l^\tau \hat{\beta} = \xi_n + \zeta_n. \]
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 $\hat{V}$ is consistent in the sense of (7), the weighted LSE $\hat{\beta}_w$ is asymptotically as efficient as $\hat{\beta}$, which is the BLUE if $V$ is known.

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

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

which is always less than 1 and equals 1 if $l^\tau \hat{\beta}$ is a BLUE (in which case $\hat{\beta} = \hat{\beta}$).

- Finding a consistent $\hat{V}$ is possible when $V$ has a certain type of structure.
Example 3.29
Consider model (1). Suppose that $V = \text{Var}(\epsilon)$ is a block diagonal matrix with the $i$th diagonal block

$$
\sigma^2 I_{m_i} + U_i \Sigma U_i^T, \quad i = 1, \ldots, k,
$$

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 $W_i$. Then

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

is an unbiased estimator of $\sigma^2$. 
Example 3.29 (continued)

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

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

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

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

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).