Chapter 3 Consistency of Least Squares Estimators

In this chapter, we continue to study the consistency of least squares estimators in various settings.

3.1 Independent Distributed Observations

Theorem 3.1.

Suppose

(i) \( y_j = x_j' \beta + e_j, \; j = 1, 2, \ldots, \beta \in \mathbb{R}^k \); 

(ii) \( \{ y_j, e_j \} \) is an independent sequence; 

(iii) \( \mathbb{E} x_j e_j = 0_x, \; j = 1, 2, \ldots \);

(iv) \( \sup\limits_{j \in \mathbb{N}} \mathbb{E} |e_j|^4 < \infty \);

(v) \( \lambda_{\min}(M_n) > \delta > 0 \) for all sufficiently large \( n \), 

where \( M_n = \frac{1}{n} \sum_{j=1}^{n} x_j x_j' \) and \( \lambda_{\min}(M_n) \) denotes the minimum eigenvalue of \( M_n \). For later reference, denote the maximum eigenvalue of matrix \( A \) by \( \lambda_{\max}(A) \).

(vi) \( \sup\limits_{j \in \mathbb{N}} \mathbb{E} |x_{jk}|^4 < \infty \), where \( x_k = (x_{1k}, \ldots, x_{nk}) \).
Then,

\[ \hat{\beta} \to \beta \quad \text{in probability}. \]

Before proving Theorem 3.1, we need some background knowledge.

**Fact 1.** If \( A \) is a \( k \times k \) symmetric matrix, then

\[ A = PDP', \]

where \( P = (P_1, \ldots, P_k) \) with \( P_i \)'s being the eigenvectors of \( A \), and \( D = \text{diag} (d_1, \ldots, d_k) \) with \( d_i \geq 0 \) being the eigenvalue corresponding to \( P_i \), that is,

\[ A P_i = d_i P_i, \quad i = 1, \ldots, k. \]

In addition, \( P'P = PP' = I \).

(Note that when \( A \) is singular, \( d_i = 0 \) for some \( i, \ldots, k \).)

**Example 1.** Let

\[ A = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \]

Then

\[ A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}. \]

\[ \lambda_{\max}(A) = \frac{3}{2}, \quad \lambda_{\min}(A) = \frac{1}{2}. \]
Example 2.

Let \( A = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \)

Then,

\[
A = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} \end{pmatrix}
\]

Q: How to explain \( P \) and \( D \) ?

Fact 2. Let \( A \) be a symmetric matrix with eigenvalues \( \lambda_1, \ldots, \lambda_k \). Then, the existence of \( \tilde{A} \) is equivalent to

Fact 3. In matrix \( A \), define its matrix norm by studying the square

\[
\| A \| = \left( \sup_{\| x \| = 1} x^T A^T A x \right)^{\frac{1}{2}}
\]

Then, it can be shown that

\[
\| A \| = \left( \lambda_{\max}(A^T A) \right)^{\frac{1}{2}}
\]

and

\[
\| \tilde{A} \| = \left( \lambda_{\min}(A^T A) \right)^{\frac{1}{2}}
\]

provided \( \tilde{A} \) exists.

In particular, if \( A \) is symmetric and nonsingular,

\[
\| A \| = \max\{ |\lambda_1|, \ldots, |\lambda_k| \}
\]

and

\[
\| \tilde{A} \| = \max\{ \frac{1}{|\lambda_1|}, \ldots, \frac{1}{|\lambda_k|} \}
\]

where \( \lambda_i \)'s are eigenvalues of \( A \).
Proof of Theorem 3.1.

It suffices to show that for any \( \varepsilon > 0 \),

\[
P( \| \hat{\beta}_n - \beta_0 \| > \varepsilon ) \to 0, \quad \text{as } n \to \infty \quad (\text{why?}) \quad (3.1)
\]

First consider the set

\[
A_n = \{ \lambda_{\min} \left( \frac{1}{n} \sum_{k=1}^{n} x_k x_k' \right) > 0 \}
\]

Then,

\[
P( ( \frac{1}{n} \sum_{k=1}^{n} x_k x_k' ) \text{ exists } )
\]

\[
= P( A_n ) \quad (\text{by Fact 2})
\]

\[
\geq P \left( \lambda_{\min} (M_n) - \| \frac{1}{n} \sum_{k=1}^{n} x_k x_k' - M_n \| > 0 \right) \quad (\text{why?})
\]

\[
= P \left( \| \frac{1}{n} \sum_{k=1}^{n} x_k x_k' - M_n \| < \lambda_{\min} (M_n) \right)
\]

\[
\geq P \left( \| \frac{1}{n} \sum_{k=1}^{n} x_k x_k' - M_n \| \leq \varepsilon \right) \quad (\text{for sufficiently large } n)
\]

\[
\to 1, \quad \text{as } n \to \infty \quad (\text{by Lemma 3.2 below}) \quad (3.2)
\]

Now,

\[
P( \| \hat{\beta}_n - \beta_0 \| > \varepsilon ) \leq P( \| \hat{\beta}_n - \beta_0 \| > \varepsilon, \ A_n ) + P( A_n^c)
\]

\[
= P( \| \hat{\beta}_n - \beta_0 \| > \varepsilon, \ A_n ) + o(1) \quad (3.3)
\]

(by (3.2))
On the set $\mathcal{A}_n$,

\[
\|\hat{\beta}_n - \beta_0\| = \left\| \left( \frac{1}{n} \sum_{k=1}^{n} x_k x_k' \right)^{-1} \frac{1}{n} \sum_{k=1}^{n} x_k \varepsilon_k \right\| 
\leq \left\| \left( \frac{1}{n} \sum_{k=1}^{n} x_k x_k' \right)^{-1} \right\| \left\| \frac{1}{n} \sum_{k=1}^{n} x_k \varepsilon_k \right\| \quad \text{(why?)} \quad (3.4)
\]

It is shown in Lemma 3.3 that for any $M > \frac{1}{\delta}$,

\[
P\left( \left\| \left( \frac{1}{n} \sum_{k=1}^{n} x_k x_k' \right)^{-1} \right\| > M, \mathcal{A}_n \right) \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (3.5)
\]

In view of (3.4) and (3.5),

\[
P\left( \|\hat{\beta}_n - \beta_0\| > \varepsilon, \mathcal{A}_n \right)
\leq P\left( \left\| \frac{1}{n} \sum_{k=1}^{n} x_k \varepsilon_k \right\| > \frac{\varepsilon}{M} \right) + o(1). \quad (3.6)
\]

According to (3.3) and (3.6), (3.1) is ensured by

\[
\|\frac{1}{n} \sum_{k=1}^{n} x_k \varepsilon_k \| = o_P(1),
\]

which is verified in Lemma 3.8.
Lemma 3.7. Assume (iii) and (vi). Then for any \( \delta > 0 \),

\[
\lim_{n \to \infty} P \left( \left| \frac{1}{n} \sum_{k=1}^{n} x_k x'_k - M_n \right| > \delta \right) = 0 \tag{3.7}
\]

where \( M_n \) is defined in (iv).

**Proof:** Define \( (U_{ij}) = E(XX') \).

Then,

\[
\frac{1}{n} \sum_{k=1}^{n} x_k x'_k - M_n = (S_{ij})
\]

with

\[
S_{ij} = \frac{1}{n} \sum_{k=1}^{n} (x_k x_{ij} - U_{ij}^{(k)})
\]

Also notice that

\[
\left| \frac{1}{n} \sum_{k=1}^{n} x_k x'_k - M_n \right| = \sqrt{\frac{\varepsilon_1 \varepsilon_2}{E_1 E_2}} S_{ij}^2 \tag{why?} \tag{3.8}
\]

Therefore,

\[
P \left( \left| \frac{1}{n} \sum_{k=1}^{n} x_k x'_k - M_n \right| > \delta \right) \leq P \left( \sqrt{\frac{\varepsilon_1 \varepsilon_2}{E_1 E_2}} S_{ij}^2 > \delta \right) \leq \frac{\frac{\varepsilon_1}{E_1} \frac{\varepsilon_2}{E_2}}{\delta} P \left( |S_{ij}| > \frac{\delta}{\sqrt{\frac{\varepsilon_1 \varepsilon_2}{E_1 E_2}}} \right) \tag{why?} \tag{3.9}
\]

In view of (3.9), (3.7) is ensured by showing that for any \( 1 \leq i, j \leq K \),

\[
P \left( |S_{ij}| > \frac{\delta}{\sqrt{\frac{\varepsilon_1 \varepsilon_2}{E_1 E_2}}} \right) = o(1). \tag{3.10}
\]
By Chebyshev's inequality,

\[ P \left( \left| \frac{1}{n} \sum_{i=1}^{\alpha} (x_i)_j - U_{ij}^{(i)} \right| > \frac{1}{k} \right) \leq \left( \frac{k}{\delta} \right)^2 \mathbb{E} \left( \left( \frac{1}{n} \sum_{i=1}^{\alpha} (x_i)_j - U_{ij}^{(i)} \right)^2 \right) \]

(by independence)

\[ \leq \left( \frac{k^2}{\delta^2} \right) \frac{4}{n^2} \sum_{i=1}^{\alpha} \left( \mathbb{E} \left( x_i^2 \right) + \mathbb{E} \left( U_{ij}^{(i)} \right)^2 \right) \]

(by (vi) and Cauchy-Schwarz's inequality)

\[ = O \left( \frac{1}{n} \right) \]

Which yields the desired result.

Lemma 3.3. Assume (iii), (v) and (vi). Then, (3.5) follows.

\[ \left\| \left( \frac{1}{n} \sum_{i=1}^{\alpha} x_i x_i^T \right)^{-1} \right\| \leq \left\| \left( \frac{1}{n} \sum_{i=1}^{\alpha} x_i x_i^T - M_n \right)^{-1} + \left\| M_n^{-1} \right\| \right\| \quad \text{why?} \]

\[ \leq \left\| \left( \frac{1}{n} \sum_{i=1}^{\alpha} x_i x_i^T - M_n \right)^{-1} \right\| \left\| \frac{1}{n} \sum_{i=1}^{\alpha} x_i x_i^T - M_n \right\| + \left\| M_n^{-1} \right\| \quad \text{why?} \]

\[ \text{Therefore,} \quad \left\| \left( \frac{1}{n} \sum_{i=1}^{\alpha} x_i x_i^T \right)^{-1} \right\| \left\| (1 - W_n) \right\| \leq \left\| M_n^{-1} \right\|. \quad (3.10) \]
By (v) and Lemma 3.2,

\[ W_n = o_p(1) \quad \text{(3.11)} \]

Now,

\[ P \left( \left\| \frac{1}{n} \sum_{j=1}^{n} X_j X_j' \right\| > M, A_n \right) \]

\[ \leq P \left( \left\| \frac{1}{n} \sum_{j=1}^{n} X_j X_j' \right\| > M, A_n, |W_n| < \frac{1}{2} \right) + P \left( |W_n| \geq \frac{1}{2} \right) \]

\[ \leq P \left( \left\| \frac{1}{n} \sum_{j=1}^{n} X_j X_j' \right\| > M (1-W_n), A_n \right) + o(1) \quad \text{(by (3.11))} \]

\[ \leq P \left( \|M_n^{-1}\| > M (1-W_n) \right) + o(1) \quad \text{(by (3.10))} \]

\[ \leq P \left( M (1-W_n) < \frac{1}{M} \right) + o(1) \quad \text{(by (v))} \]

for all sufficiently large \( n \)

\[ = P \left( W_n > \frac{M - \frac{1}{M}}{M} \right) \]

\[ = o(1). \quad \text{(by (3.11) and the hypothesis that } M > \frac{1}{2} \text{)} \]

Consequently, (3.5) holds.
Lemma 3.4. Assume (ii), (iii), (iv) and
\[ \sup_{i \in \mathbb{N}} \mathbb{E} X_i^2 < \infty. \] Then,
\[ \left( \frac{1}{n} \sum_{i=1}^{n} X_i : X_i \right) \xrightarrow{p} 0 \quad (3.2) \]

(Proof) Notice that (3.2) is guaranteed by
for all \( i \leq k \),
\[ \frac{1}{n} \sum_{i}^{n} X_i : X_i = o_p(1). \] (why?) (3.3)

Now, for any \( \varepsilon > 0 \),
\[ P \left( \left| \frac{1}{n} \sum_{i=1}^{n} X_i : X_i \right| > \varepsilon \right) \]
\[ \leq \frac{1}{\varepsilon^2} \frac{1}{n^2} \sum_{i}^{n} \mathbb{E} \left( X_i^2 \right) \] (by (ii), (iii) and Chebyshev's inequality)
\[ \leq \frac{1}{\varepsilon^2} \frac{1}{n} \sup_{i \geq 0} \mathbb{E} \mathbb{X}_i^4 \sup_{i \geq 0} \mathbb{E} \mathbb{X}_i^4 \] (by Cauchy-Schwarz's inequality)
\[ = o(1), \] (by (iv) and (vi))

which gives (3.3).
Remark:

For one of the most beautiful results on the consistency of least squares estimators, see Lai and Wei (1982), Least squares estimator in stochastic regression models with applications to identification and control of dynamic systems, *Annals of Statistics*, 1982, 10, 154-166.

Exercise 1 (Bonus Problem)

Let \{Z_t\} be a sequence of independent random variables with \(E|Z_t| < \infty\) for all \(t\). Then, by Burkholder's inequality, it can be shown that for \(p > 0\), there is a positive constant \(C_0\) such that

\[
E \left( \sum_{t=1}^{n} |Z_t| \right)^p \leq C_0 \left( \sum_{t=1}^{n} E|Z_t|^p \right)
\]

Use (*) to show that (iv) and (vi) in Theorem 3.1 can be weakened to

(iv') \(\sup_{1 \leq k < \infty} E|X_k| < \infty\)

and (vi') \(\sup_{1 \leq k < \infty} E|X_k| < \infty\), respectively, where \(0 < c < 1\) is some positive constant.
Exercise 2. Consider IV estimator \( \beta_n = (X' \hat{Z} Z X)^{-1} X' \hat{Z} Y \).

Suppose

(i) \( Y_t = X_t' \beta_0 + \epsilon_t, \quad t = 1, 2, \ldots, \beta \in \mathbb{R}^k \);

(ii) \{ X_t, Z_t, \epsilon_t \} \) is an independent sequence;

(iii) \( E \epsilon_t Z_t = 0 \), where \( k \leq K \);

(iv) \( \sup_{|t| \leq n} E |\epsilon_t|^4 < \infty \);

(v) \( \lambda_{\min}(\Omega_n \Omega_n) > \delta_1 > 0 \) for all sufficiently large \( n \), where \( \Omega_n = \frac{1}{n} E(\sum_{t=1}^{n} Z_t X_t' ) \).

(vi) \( \hat{\beta}_n - \beta_0 \xrightarrow{\text{prob.}} 0 \), where \( \Omega_n \) is a symmetric matrix satisfying \( \lambda_{\min}(\Omega_n) > \delta_2 > 0 \) for all sufficiently large \( n \) and \( \lambda_{\max}(\Omega_n) < \bar{M} < \infty \).

(vii) \( \sup_{|t| \leq n} E |X_t|^4 < \infty \), \( t = 1, \ldots, K \).

(viii) \( \sup_{|t| \leq n} E |Z_{tm}|^4 < \infty \), \( m = 1, \ldots, k \).

Then, \( \hat{\beta}_n \rightarrow \beta_0 \) in probability.