

Linear Regression with independent data

Consider the usual regression model

$$y_i = \underbrace{x_i' \beta}_{k \times 1} + \underbrace{u_i}_{k \times 1}$$

Identification

(i) $E x_i u_i = 0$

(ii) $E x_i x_i'$ has full rank k .

Then $E x_i (y_i - x_i' \beta_0) = 0$

$$\Rightarrow E x_i y_i = E x_i x_i' \beta$$

$$\Rightarrow \beta = (E x_i x_i')^{-1} E x_i y_i$$

The proposed estimator is the sample analogue :

$$\hat{\beta}_n = \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i y_i$$

Asymptotic Theory

Lemma 1 :- (from Portmanteau Theorem) Convergence in distribution of random vectors in \mathbb{R}^k , $x_n \xrightarrow{d} x$, is equivalent to

$$\lim_{n \rightarrow \infty} E h(x_n) = E h(x) \quad \text{for every continuous functional } h(\cdot).$$

(Intuition : CDF converge if all the moments converge)

Lemma 2 :- a) (Monotone Convergence) If $x_n \geq 0$ and for each $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} x_n(\omega) = x(\omega). \text{ Then}$$

$$\lim_{n \rightarrow \infty} E x_n = E x$$

b) (Fatou's Lemma) If $x_n \geq 0$, then

$$E (\liminf_{n \rightarrow \infty} x_n) \leq \liminf_{n \rightarrow \infty} E x$$

c) (Dominated Convergence) Suppose that for each $w \in \Omega$,
 $\lim_{n \rightarrow \infty} E X_n = EX$. Furthermore, there is some Y such that
 $E Y < \infty$ and $|X_n| \leq Y$ for each $n \geq 1$. Then

$$\lim_{n \rightarrow \infty} E X_n = EX$$

Lemma 3. - We say

$$X_n \xrightarrow{d} X \iff \begin{array}{l} \text{for all open } G \subset \mathbb{R}^d, \\ \liminf_{n \rightarrow \infty} P\{X_n \in G\} \geq P\{X \in G\} \end{array}$$

proof: Consider sufficiency first. We assume that B holds. Then take some continuous function $f(\cdot)$ and write

$$\begin{aligned} E f(X) &= \int_0^\infty P\{f(X) > t\} dt \\ &\stackrel{\text{by B}}{\leq} \int_0^\infty \liminf_{n \rightarrow \infty} P\{f(X_n) > t\} dt \\ &\stackrel{\substack{\text{by dominated convergence} \\ \text{with dominating function} \\ Y = 1.}}{\leq} \liminf_{n \rightarrow \infty} \int_0^\infty P\{f(X_n) > t\} dt \\ &= \liminf_{n \rightarrow \infty} E f(X_n) \end{aligned}$$

Now take some bounded continuous function $h(\cdot)$ such that $|h| \leq c < \infty$. This requires that $c + h(\cdot) \geq 0$ and $c - h(\cdot) \geq 0$. Then

$$\begin{aligned} \bullet \quad c + E h(X) &\leq c + \liminf_{n \rightarrow \infty} E(X_n) \\ \Rightarrow \quad E h(X) &\leq \liminf_{n \rightarrow \infty} E(X_n) \\ \bullet \quad c - E h(X) &\leq c - \liminf_{n \rightarrow \infty} E(X_n) \\ \Rightarrow \quad E h(X) &\geq \limsup_{n \rightarrow \infty} E(X_n) \end{aligned}$$

So we conclude that $\lim_{n \rightarrow \infty} E h(X_n) = E h(X)$. By Lemma 2 our result follows.

For necessity take a continuous nonnegative $f_j(X) \leq \mathbb{1}_{\{X \in G\}}$ such that $f_j(\cdot) \rightarrow \mathbb{1}_{\{\cdot \in G\}}$ as $j \rightarrow \infty$. Then the result follows from Fatou's Lemma. ■

Theorem 2. - (Continuous Mapping) Let $\{x_n\}$ be a sequence of random vectors in \mathbb{R}^d such that $x_n \xrightarrow{P} x$. Also, let $g(\cdot)$ be an \mathbb{R}^k -valued function on \mathbb{R}^d that is continuous on a set C such that $P\{x \in C\} = 1$ (i.e. almost everywhere). Then

$$g(x_n) \xrightarrow{P} g(x) \quad \text{as } n \rightarrow \infty.$$

proof: Assume $x_n \xrightarrow{P} x$. We want to show $P\{g(x) \in G\} \leq \liminf_{n \rightarrow \infty} P\{g(x_n) \in G\}$ where G is an open set in \mathbb{R}^d .

Suppose $g^{-1}(G) \cap C$ is empty. Then

$$\begin{aligned} P\{g(x) \in G\} &= P[\{g(x) \in G\} \cap \{x \in C\}] \\ &= P[x \in g^{-1}(G) \cap C] = 0 \\ &\leq \liminf_{n \rightarrow \infty} P\{g(x_n) \in G\} \text{ is trivially true.} \end{aligned}$$

Suppose $g^{-1}(G) \cap C$ is non-empty. Then choose a point $v \in g^{-1}(G) \cap C$. There must be an open ball N_v such that $g(w) \in G$ for all $w \in N_v$. This means that $N_v \subset g^{-1}(G)$. Thus $g^{-1}(G) \cap C = \underbrace{g^{-1}(G)}_{\text{interior of } g^{-1}(G)}$

Using this result we write

$$\begin{aligned} P\{\bar{x} \in g^{-1}(G)\} &= P\{x \in g^{-1}(G) \cap C\} \\ &\leq P\{x \in g^{-1}(G)^{\circ}\} \\ &\leq \liminf_{n \rightarrow \infty} \{x_n \in g^{-1}(G)^{\circ}\} \\ &\leq \liminf_{n \rightarrow \infty} \{x_n \in g^{-1}(G)\}. \end{aligned}$$

by Lemma 1, the result follows.

Lemma 4. - (Slutsky's Theorem) Suppose that a sequence of random vectors $x_n \xrightarrow{P} x$ and let $h(\cdot)$ be continuous almost everywhere. Then $h(x_n) \xrightarrow{P} h(x)$.

proof: We want to show that $\limsup_{n \rightarrow \infty} P\{|h(x_n) - h(x)| > \epsilon\} = 0$. Write

$$\begin{aligned} P(|h(x_n) - h(x)| > \epsilon) &= P(|h(x_n) - h(x)| > \epsilon, \|x_n - x\| \geq f_\epsilon) + \\ &\quad P(|h(x_n) - h(x)| > \epsilon, \|x_n - x\| < f_\epsilon) \end{aligned}$$

because continuity implies that the second term is zero

$$\begin{aligned}
 &= P(\|h(x_n) - h(x)\| > \varepsilon, \|x_n - x\| \geq \delta_\varepsilon) \\
 &\leq P(\|x_n - x\| \geq \delta_\varepsilon) \\
 &= P(o_p(1) \geq \delta_\varepsilon)
 \end{aligned}$$

Taking limit yields

$$\limsup_{n \rightarrow \infty} P(\|h(x_n) - h(x)\| > \varepsilon) \leq \limsup_{n \rightarrow \infty} P(o_p(1) \geq \delta_\varepsilon) = 0.$$

Theorem 2 :- (iid Weak Law of Large Numbers) Let $\{X_n\}$ be a sequence of iid random vectors such that $E\|X_i\| < \infty$. Then

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} EX_i$$

proof: Denote the characteristic function of X_i as $\varphi(t) := E[\exp(jt X_i)]$, where $j = \sqrt{-1}$. Write

$$\begin{aligned}
 \varphi_n(t) &= E \left[\exp \left(jt \sum_{i=1}^n X_i \right) \right] \\
 &= E \left[\prod_{i=1}^n \exp \left(jt \frac{X_i}{n} \right) \right] \\
 \text{by indep} &= \prod_{i=1}^n E \left[\exp \left(jt \frac{X_i}{n} \right) \right] \\
 \text{by id} &= \left\{ E \left[\exp \left(jt \frac{X_i}{n} \right) \right] \right\}^n \\
 &= \{\varphi(t)\}^n.
 \end{aligned}$$

By a Taylor approximation around $t=0$:

$$\varphi(t) = \underbrace{\varphi(0)}_{=1} + t' E[j X_i] + o\left(\frac{1}{n}\right)$$

Then

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \varphi_n(t) &= \lim_{n \rightarrow \infty} \left\{ 1 + jt' E[X_i] + o\left(\frac{1}{n}\right) \right\}^n \\
 &= \exp(jt' E[X_i]).
 \end{aligned}$$

By Levy's continuity theorem we conclude that

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{d} E[X_i] \quad \text{which implied convergence in probability because it's a degenerate distribution.}$$

Theorem 3. - (i.i.d Weak Law of Large Numbers) Suppose that $\{X_n\}$ is a sequence of mean zero random vectors in \mathbb{R}^k such that $E X_i, E X_j, m = 0$ for all $i \neq j$, $i \in \{1, 2, \dots, k\}$ and $m \in \{1, 2, \dots, k\}$. Moreover,

$$\frac{1}{n} \max_{1 \leq j \leq n} E \|X_j\|^2 = o(1)$$

then

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} 0.$$

proof: Write

$$\begin{aligned} P\left(\left\|\frac{1}{n} \sum_{i=1}^n X_i\right\| > \varepsilon\right) &\stackrel{\text{Markov ineq}}{\leq} E\left[\left\|\frac{1}{n} \sum_{i=1}^n X_i\right\|^2\right] \frac{1}{\varepsilon^2} \\ &= E\left[\frac{1}{n^2} \sum_{i=1}^n \|X_i\|^2\right] \frac{1}{\varepsilon^2} \\ &\leq \frac{1}{n^2} n \max_{1 \leq j \leq n} E \|X_j\|^2 \\ &= o(1) \end{aligned}$$

Then the result follows from taking limit as $n \rightarrow \infty$ on both sides.

Theorem 4. - (i.i.d Central Limit) Suppose $\{X_n\}$ are a sequence of i.i.d random vectors such that $E \|X_i\|^2 < \infty$ and $\text{Var}(X_i) = \Sigma$ and non singular. Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - E X_i) \xrightarrow{d} N(0, \Sigma)$$

proof:

We can use Levy's continuity theorem for the univariate case and use the Cramer-Wold device to extend it to random vectors. I'll omit the details of this proof.

Consistency

- (i) Data $\{y_i, x_i'\}$ is iid
- (ii) $E x_i u_i = 0$
- (iii) $E x_i x_i'$ has full rank n .

Then

$$\hat{\beta}_n = \beta + o_p(1)$$

Proof:

$$\begin{aligned}
 \hat{\beta}_n &= \beta + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i u_i \\
 &= \beta + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' - E x_i x_i' + E x_i x_i' \right)^{-1} \underbrace{\frac{1}{n} \sum_{i=1}^n (x_i u_i + E x_i u_i - E x_i u_i)}_{=0} \\
 &\stackrel{\text{Theorem 3}}{=} \beta + \left[o_p(1) + E x_i x_i' \right]^{-1} \left[o_p(1) + \underbrace{E x_i u_i}_{=0} \right] \\
 &\stackrel{\text{Lemma 4}}{=} \beta + \left[o_p(1) + (E x_i x_i')^{-1} \right] o_p(1) \\
 &= \beta + o_p(1). \quad \blacksquare
 \end{aligned}$$

Asymptotic Normality

- (i) Data $\{y_i, x_i'\}$ is iid
- (ii) $E x_i u_i = 0$
- (iii) $E x_i x_i'$ has full rank n .
- (iv) $E x_{ij}^4 < \infty$ for $j=1, \dots, k$.
- (v) $E u_i^4 < \infty$
- (vi) $\text{Var}(x_i u_i)$ is positive definite

Lemma 5. - Provided (iv) - (vi) hold. Then $\text{Var}(x_i u_i) = O(1)$.

Proof: $\text{Var}(x_i u_i) = E(u_i^2 x_i x_i')$

$$\begin{aligned}
& \stackrel{\text{Cauchy-Schwarz}}{\leq} \{E |u_i|^4 E \|x_i x_i'\|^2\}^{1/2} \\
& = \{E |u_i|^4 E (\text{tr}(x_i x_i' x_i x_i'))\}^{1/2} \\
& \stackrel{\text{Frobenius Norm}}{=} \{E |u_i|^4 E (x_i' x_i)^2\}^{1/2} \\
& \text{Recall } \|x_i\| = (x_i' x_i)^{1/2} \\
& = \{E |u_i|^4 E \|x_i\|^4\}^{1/2} \\
& = \{E |u_i|^4 E [\sum_{j=1}^n x_{ij}^2]^2\}^{1/2} \\
& \text{for some constant } C \leq \{E |u_i|^4 C \max_{1 \leq j \leq n} E [x_{ij}^4]\}^{1/2} \\
& = O(1) \cdot O(1) \\
& = O(1) \quad .
\end{aligned}$$

Now assume all conditions (i)-(vi) hold. Then

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, (E x_i x_i')^{-1} \text{Var}(x_i u_i) (E x_i x_i')^{-1})$$

proof:

First, notice that with Lemma 5 and Theorem 4 we have

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i \xrightarrow{d} N(0, \text{Var}(x_i u_i)) \\
& \text{i.e. } \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i = O_p(1).
\end{aligned}$$

Then,

$$(E x_i x_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i \xrightarrow{d} N(0, (E x_i x_i')^{-1} \text{Var}(x_i u_i) (E x_i x_i')^{-1}). \quad ■$$

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_n - \beta) &= \left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i \\
&= \left[(E x_i x_i')^{-1} + o_p(1) \right] \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i
\end{aligned}$$

$$\begin{aligned}
&= (\mathbb{E} \mathbf{x}_i \mathbf{x}_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i + o_p(1) \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i \\
&\stackrel{\text{Theorem 4}}{=} (\mathbb{E} \mathbf{x}_i \mathbf{x}_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i + o_p(1) O_p(1) \\
&= (\mathbb{E} \mathbf{x}_i \mathbf{x}_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i + o_p(1). \\
&\xrightarrow{d} \mathbb{E}(\mathbf{x}_i \mathbf{x}_i')^{-1} N(0, \text{Var}(x_i u_i)) \\
&= N(0, (\mathbb{E} \mathbf{x}_i \mathbf{x}_i')^{-1} \text{Var}(x_i u_i) \mathbb{E}(\mathbf{x}_i \mathbf{x}_i')^{-1}) \\
&\equiv N(0, V).
\end{aligned}$$

Estimation of Asymptotic Variance Matrix

Let $\hat{M}_n = \frac{1}{n} \sum \mathbf{x}_i \mathbf{x}_i'$, $\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i'$, $\hat{u}_i = y_i - \mathbf{x}_i' \hat{\beta}_n$.

We propose the following estimator of the asymptotic variance V :

$$\hat{V}_n = \hat{M}_n^{-1} \hat{\Omega}_n \hat{M}_n^{-1}.$$

First, consider

$$\begin{aligned}
\hat{\Omega}_n &= \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i' \\
&= \frac{1}{n} \sum_{i=1}^n [(\mathbf{y}_i - \mathbf{x}_i' \hat{\beta}_n \pm \mathbf{x}_i' \beta)^2 \mathbf{x}_i \mathbf{x}_i'] \\
&= \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i' + \frac{1}{n} \sum_{i=1}^n [\mathbf{x}_i' (\hat{\beta}_n - \beta)]^2 \mathbf{x}_i \mathbf{x}_i' - \frac{2}{n} \sum_{i=1}^n [\mathbf{x}_i' (\hat{\beta}_n - \beta) \hat{u}_i] \mathbf{x}_i \mathbf{x}_i' \\
&\stackrel{\text{Theorem 4}}{=} \mathcal{R} + o_p(1) + R_{1n} + R_{2n}
\end{aligned}$$

$$\begin{aligned}
\|\mathcal{R}_{2n}\| &\leq \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_i' (\hat{\beta}_n - \beta)|^2 \|\mathbf{x}_i\|^2 \\
&\leq \|\hat{\beta}_n - \beta\|^2 \frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{x}_i\|^2 \\
&= o_p(1) \frac{1}{n} \sum_{j=1}^n \|\mathbf{x}_i\|^4
\end{aligned}$$

$$\begin{aligned}
&= o_p(1) \left[E \|x_i\|^4 + o_p(1) \right] \\
&= o_p(1) [O(1) + o_p(1)] \\
&= o_p(1).
\end{aligned}$$

$$\begin{aligned}
\bullet \|R_{2n}\| &\leq \|\beta_n - \beta\| \frac{1}{n} \sum_{i=1}^n |u_i| \|x_i\| \|x_i x_i'\| \\
&\leq \|\beta_n - \beta\| \frac{1}{n} \sum_{i=1}^n |u_i| \|x_i\|^3 \\
&= o_p(1) \left[E |u_i| \|x_i\|^3 + o_p(1) \right] \\
&\leq o_p(1) \left[(E |u_i|^2 \|x_i\|^2)^{1/2} (E \|x_i\|^4)^{1/2} + o_p(1) \right] \\
&\leq o_p(1) \left[(\sqrt{E |u_i|^4 E \|x_i\|^4})^{1/2} (E \|x_i\|^4)^{1/2} + o_p(1) \right] \\
&= o_p(1) [O(1) O(1) + o_p(1)] \\
&= o_p(1).
\end{aligned}$$

Putting it all together yields

$$\hat{\Sigma}_n = \Sigma + o_p(1).$$

Finally, write

$$\begin{aligned}
\hat{\Sigma}_n &= M_n^{-1} \hat{\Sigma}_n M_n^{-1} \\
&= [(E x_i x_i')^{-1} + o_p(1)] [\Sigma + o_p(1)] [(E x_i x_i')^{-1} + o_p(1)] \\
&= (E x_i x_i')^{-1} \Sigma E (x_i x_i')^{-1} + o_p(1) \\
&\quad \text{because } (E x_i x_i')^{-1} = O(1) \\
\Sigma &= O(1) = V + o_p(1). \quad \blacksquare
\end{aligned}$$