Linear Regression with independent data

Consider the uwoe regression model

$$\underset{1\times1}{y_i} = \underset{1\times k}{X_i'}\underset{k\times1}{\beta} + \underset{1\times1}{u_i}$$

Identification

(i) $E\, X_i u_i = 0$

(ii) $E\, X_i X_i'$ has full rank k.

Then $$E\, X_i (y_i - X_i'\beta) = 0$$

$$\Rightarrow E\, X_i y_i = E\, X_i X_i' \beta$$

$$\Rightarrow \boxed{\beta = (E\, X_i X_i')^{-1} E\, X_i y_i}$$

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

$$\lim_{n\to\infty} E\, h(X_n) = E\, h(X)$$ for every continous functional $h(\cdot)$.

(Intuition : CDF converge if all the moments converge)

Lemma 2 :- a) (Monotone Convergence) If $X_n \geq 0$, monotonic and for each $\omega \in \Omega$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Then

$$\lim_{n\to\infty} E X_n = E X$$

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

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} E X_n$$

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

$$\lim_{n\to\infty} E X_n = E X$$

<u>Lemma 3</u>:- We say

$$\underbrace{X_n \xrightarrow[d]{} X}_{A} \iff \text{for all open } G \subset \mathbb{R}^d, \ \underbrace{\liminf_{n\to\infty} P\{X_n \in G\} \geq P\{X \in G\}}_{B}$$

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

$$E f(X) = \int_0^\infty P\{f(X) > t\}\, dt$$

$$\underset{\text{by } B}{\leq} \int_0^\infty \liminf_{n\to\infty} P\{f(X_n) > t\}\, dt$$

$$\underset{\text{Fatou's Lemma}}{\leq} \liminf_{n\to\infty} \int_0^\infty P\{f(X_n) > t\}\, dt$$

$$= \liminf_{n\to\infty} E f(X_n)$$

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

- $$c + E h(X) \leq c + \liminf_{n\to\infty} E(X_n)$$

$$\Rightarrow E h(X) \leq \liminf_{n\to\infty} E(X_n)$$

- $$c - E h(X) \leq c - \liminf_{n\to\infty} E(X_n)$$

$$\Rightarrow E h(X) \geq \limsup_{n\to\infty} E(X_n)$$

So we conclude that $\lim_{n\to\infty} E h(X_n) = E h(X)$. By Lemma 1 our result follows.

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

Theorem 1.- (Continuous Mapping) Let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^d$ such that $X_n \xrightarrow{d} X$. 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

$$g(X_n) \xrightarrow{d} g(X) \quad \text{as} \quad n \to \infty.$$

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

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

$$P\{g(X) \in G\} = P[\{g(X) \in G\} \cap \{X \in C\}]$$

$$= P[X \in g^{-1}(G) \cap C] = 0$$

$$\leq \liminf_{n\to\infty} P\{g(X_n) \in G\} \text{ is trivially true.}$$

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 $g(\omega) \in G$ for all $\omega \in N_v$. This means that $N_v \subset g^{-1}(G)$. Thus $g^{-1}(G) \cap C \subset \underbrace{g^{-1}(G)^\circ}_{\text{interior of } g^{-1}(G)}$

Using this result we write

$$P\{X \in g^{-1}(G)\} = P\{X \in g^{-1}(G) \cap C\}$$

$$\leq P\{X \in g^{-1}(G)^\circ\}$$

because $X_n \xrightarrow{d} X$

$$\leq \liminf_{n\to\infty} P\{X_n \in g^{-1}(G)^\circ\}$$

$$\leq \liminf_{n\to\infty} P\{X_n \in g^{-1}(G)\}.$$

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 continous almost everywhere. Then

$$h(X_n) \xrightarrow{P} h(X).$$

proof: We want to show that $\limsup_{n\to\infty} P\{\|h(X_n) - h(X)\| > \varepsilon\} = 0$. write

$$P(\|h(X_n) - h(X)\| > \varepsilon) = P(\|h(X_n) - h(X)\| > \varepsilon, \|X_n - X\| \geq \delta_\varepsilon) +$$
$$P(\|h(X_n) - h(X)\| > \varepsilon, \|X_n - X\| < \delta_\varepsilon)$$

because continuity implies that the second term is zero

$$= P(\|h(X_n) - h(X)\| > \varepsilon,\ \|X_n - X\| \geq \delta_\varepsilon)$$

$$\leq P(\|X_n - X\| \geq \delta_\varepsilon)$$

$$= P(o_p(1) \geq \delta_\varepsilon)$$

Taking limit yields

$$\limsup_{n\to\infty} P(\|h(X_n) - h(X)\| > \varepsilon) \leq \limsup_{n\to\infty} P(o_p(1) \geq \delta_\varepsilon)$$

$$= 0. \quad \blacksquare$$

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

$$\varphi_n(t) = E\left[\exp\left(\frac{jt}{n}\sum_{i=1}^n X_i\right)\right]$$

$$= E\left[\prod_{i=1}^n \exp\left(\frac{jt}{n} X_i\right)\right]$$

by indep $$= \prod_{i=1}^n E\left[\exp\left(\frac{jt}{n} X_i\right)\right]$$

by id $$= \left\{E\left[\exp\left(\frac{jt}{n} X_i\right)\right]\right\}^n$$

$$= \{\varphi(t/n)\}^n.$$

By a Taylor approximation around $t=0$:

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

Then

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} \left\{1 + j\frac{t'}{n} E[X_i] + o\left(\frac{1}{n}\right)\right\}^n$$

$$= \exp(jt' E[X_i]).$$

By Levy's continuity theorem we conclude that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[d]{} E[X_i]$$

which implies convergence in probability because it's a degenerate distribution. ∎

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

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

Then

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

proof: Write

$$P\left(\left\| \frac{1}{n}\sum_{i=1}^{n} X_i \right\| > \varepsilon \right) \underset{\text{Markov ineq}}{\le} E\left[\left\| \frac{1}{n}\sum_{i=1}^{n} X_i \right\|^2 \right] \frac{1}{\varepsilon^2}$$

$$\underset{\text{by indep}}{=} E\left[\frac{1}{n^2}\sum_{i=1}^{n} \|X_i\|^2 \right] \frac{1}{\varepsilon^2}$$

$$\le \frac{1}{n^2}\, n \max_{1\le j\le n} E\,\|X_j\|^2$$

$$= o(1)$$

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

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

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

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 ommit 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 k

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

proof:

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

$$= \beta + \Big(\underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i x_i' - E x_i x_i'}_{} + E x_i x_i' \Big)^{-1} \underbrace{\frac{1}{n} \sum_{i=1}^{n} (x_i u_i + E x_i u_i - E x_i u_i}_{})$$

Theorem 3
$$= \beta + \left(o_p(1) + E x_i x_i' \right)^{-1} \Big[o_p(1) + \underbrace{E x_i u_i}_{=0} \Big]$$

Lemma 4
$$= \beta + \left[o_p(1) + (E x_i x_i')^{-1} \right] o_p(1)$$

$$= \beta + o_p(1) . \quad \blacksquare$$

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 k.

(iv) $E\, x_{ij}^4 < \infty$ for $j = 1, \dots, k$.

(v) $E\, u_i^4 < \infty$

(vi) $Var(x_i u_i)$ is positive definite

Lemma 5.- Provided (iv) - (v) hold. Then $Var(x_i u_i) = O(1)$.

proof: $Var(x_i u_i) = E(u_i^2 x_i x_i')$

Cauchy Schwarz $\leq \{E\,|u_i|^4 \; E\,\|x_i x_i'\|^2\}^{1/2}$

$= \{E\,|u_i|^4 \; E(tr(x_i x_i' \, x_i x_i'))\}^{1/2}$

Frobenius Norm

$= \{E\,|u_i|^4 \; E(x_i' x_i)^2\}^{1/2}$

Recall $\|x_i\| = (x_i' x_i)^{1/2}$

$= \{E\,|u_i|^4 \; E\,\|x_i\|^4\}^{1/2}$

$= \{E\,|u_i|^4 \; E[\sum_{j=1}^{k} x_{ij}^2]^2\}^{1/2}$

for some constant c $\leq \{E\,|u_i|^4 \, C \max_{1 \leq j \leq k} E[x_{ij}^4]\}^{1/2}$

$= O(1) \,.\, O(1)$

$= O(1)$. ∎

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

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

proof:

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i \xrightarrow[d]{} N(0, Var(x_i u_i))$$

i.e. $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i = O_p(1)$.

Then, we write

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

$$= \left[(E x_i x_i')^{-1} + o_p(1)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i$$

$$= (E x_i x_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i + o_p(1) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i$$

Theorem 4

$$= (E x_i x_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i + o_p(1)\, O_p(1)$$

$$= (E x_i x_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i + o_p(1) .$$

$$\xrightarrow[d]{} E(x_i x_i')^{-1} N(0, Var(x_i u_i))$$

$$= N(0, (E x_i x_i')^{-1} Var(x_i u_i) E(x_i x_i')^{-1})$$

$$\equiv N(0, V) . \quad \blacksquare$$

Estimation of Asymptotic Variance Matrix

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

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

$$\Omega_n = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i^2 x_i x_i'$$

$$= \frac{1}{n} \sum_{i=1}^{n} [(y_i - x_i' \hat{\beta}_n \pm x_i' \beta)^2 x_i x_i']$$

$$= \frac{1}{n} \sum_{i=1}^{n} u_i^2 x_i x_i' + \frac{1}{n} \sum_{i=1}^{n} [x_i' (\hat{\beta}_n - \beta)]^2 x_i x_i' - \frac{2}{2} \sum_{i=1}^{n} [x_i' (\hat{\beta}_n - \beta) u_i] x_i x_i'$$

$$= \Omega + o_p(1) + R_{1n} + R_{2n}$$

Theorem 4

- $$\| R_{1n} \| \leq \frac{1}{n} \sum_{j=1}^{n} | x_i' (\hat{\beta}_n - \beta) |^2 \| x_i x_i' \|$$

$$\leq \| \hat{\beta}_n - \beta \|^2 \frac{1}{n} \sum_{j=1}^{n} \| x_i \|^2 \| x_i \|^2$$

$$= o_p(1) \frac{1}{n} \sum_{j=1}^{n} \| x_i \|^4$$

$$= o_p(1)\left[E\|X_i\|^4 + o_p(1)\right]$$

$$= o_p(1)\left[O(1) + o_p(1)\right]$$

$$= o_p(1).$$

- $\|R_{2n}\| \leq \|\hat{\beta}_n - \beta\| \frac{1}{n}\sum_{i=1}^{n} |u_i| \, \|X_i\| \, \|X_iX_i'\|$

$$\leq \|\hat{\beta}_n - \beta\| \frac{1}{n}\sum_{i=1}^{n} |u_i| \, \|X_i\|^3$$

$$= o_p(1)\left[E|u_i|\,\|X_i\|^3 + o_p(1)\right]$$

$$\leq o_p(1)\left[\left(E|u_i|^2\|X_i\|^2\right)^{1/2} \left(E\|X_i\|^4\right)^{1/2} + o_p(1)\right]$$

$$\leq o_p(1)\left[\left(\sqrt{E|u_i|^4 \; E\|X_i\|^4}\right)^{1/2} \left(E\|X_i\|^4\right)^{1/2} + o_p(1)\right]$$

$$= o_p(1)\left[O(1)\,O(1) + o_p(1)\right]$$

$$= o_p(1).$$

Putting it all together yields

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

Finally, write

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

$$= \left[(EX_iX_i')^{-1} + o_p(1)\right]\left[\Omega + o_p(1)\right]\left[(EX_iX_i')^{-1} + o_p(1)\right]$$

$$= (EX_iX_i')^{-1}\,\Omega\, E(X_iX_i')^{-1} + o_p(1)$$

because $(EX_iX_i')^{-1} = O(1)$, $\Omega = O(1)$

$$= V + o_p(1). \quad \blacksquare$$