Rozszerzenia KMRL

MNK z losową macierzą obserwacji

- Równanie modelu

$$y = X\beta + \varepsilon$$

- Jeżeli X zawiera elementy losowe to należy sprawdzić czy

$$E(b-\beta) = E[(X'X)^{-1}X'\varepsilon] \stackrel{?}{=} E[(X'X)^{-1}X']E(\varepsilon)$$

- Przypomnienie:

Nieskorelowane zmienne losowe

Z teorii RP wiemy, że jeśli X, Y zmienne losowe

$$E(XY) = E(X)E(Y) \Leftrightarrow cov(X, Y) = 0$$

Zmodyfikowane założenia

1. Proces generujący dane,
2. Liniowość,
3. Warunkowa wartość oczekiwana składnika losowego jest równa zero,

$$\forall_i \; E(\varepsilon_i \mid X) = 0$$

4. Warunkowa wariancja składnika losowego jest identyczna dla każdej obserwacji,

$$\forall_i \; var(\varepsilon_i \mid X) = \sigma^2$$

Wnioski 1/2

- Konsekwencją 3 założenia jest

$$E(y \mid X) = E(X\beta + \varepsilon \mid X) = E(X\beta \mid X) + \underbrace{E(\varepsilon \mid X)}_{0} = X\beta$$

- Warunkowa wartość oczekiwana estymatora

$$E(b \mid X) = \beta + E((X'X)^{-1}X'\varepsilon \mid X) = \beta + (X'X)^{-1}X'\underbrace{E(\varepsilon \mid X)}_{0} = \beta$$

- Estymator MNK jest również bezwarunkowo nieobciążony

$$E(E(b \mid X)) = E(\beta) = \beta$$

Wnioski 2/2

- Podobnie uzyskiwany jest wzór na warunkową wariancję estymatora

$$var(b \mid X) = var(\beta + (X'X)^{-1}X'\varepsilon \mid X) =$$

$$= (X'X)^{-1}X' var(\varepsilon \mid X) X(X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

- Rozkłady estymatorów w próbach o małej liczbie obserwacji wyprowadzane są przy założeniu

$$\varepsilon \mid X \ \sim \ \mathcal{N}(0, \sigma^2 \mathbb{I})$$

Zgodność estymatora

1. $E(x_i\varepsilon_i) = 0 \quad \forall i$
2. $E(x_i'x_i) = \Sigma_x$ i jest odwracalna

Zbieżność według prawdopodobieństwa

Ciąg zmiennych losowych a_n jest zbieżny według prawdopodobieństwa do granicy a gdy

$$\forall\ \delta > 0 \ \lim_{n\to\infty} Pr(| \ a_n - a \ |> \delta) = 0$$

alternatywnie

$$\forall\ \delta > 0 \ \lim_{n\to\infty} Pr(| \ a_n - a \ |\leq \delta) = 1$$

Zbieżność według prawdopodobieństwa oznaczamy symbolem plim

Przykład. Zbieżność według prawdopodobieństwa

- Rozpatrzmy ciąg rzutów prawidłową symetryczną monetą. Niech X_n będzie zdefiniowane

$$X_n = \begin{cases} 1 & \text{orzeł} \\ 0 & \text{reszka} \end{cases}$$

- Niech $\frac{S_n}{n}$ będzie średnią wartością zmiennej losowej z n rzutów
- Zatem dla każdej $\delta > 0$

$$\lim_{n \to \infty} Pr\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \delta\right) = 0$$

- Ponieważ średnia zbiega według prawdopodobieństwa do $\frac{1}{2}$

Zgodność estymatora

3. $\text{plim}(\frac{1}{N}\sum x_i'\varepsilon_i) = E(x_i'\varepsilon_i)$
4. $\text{plim}(\frac{1}{N}\sum x_i'x_i) = E(x_i'x_i)$

Twierdzenie Słuckiego (1925)

Jeżeli funkcja $g(\cdot)$ jest ciągła w otoczeniu punktu a, to

$$a_n \xrightarrow{p} a \Longrightarrow g(a_n) \xrightarrow{p} g(a)$$

zapis alternatywny

$$\text{plim}\, g(a_n) = \text{plim}\, g(a)$$

Wnioski z Twierdzenia Słuckiego

- Podstawowa zaleta

$$E(g(x_n)) \neq g(E(x_n)) \ \textit{ale} \ \mathrm{plim}(g(x_n)) = g(\mathrm{plim}(x_n))$$

- Wnioski. Jeżeli $a_n \xrightarrow{p} a$, $b_n \xrightarrow{p} b$, $C_n \xrightarrow{p} C$
  1. $\mathrm{plim}(a_n \pm b_n) = a \pm b$
  2. $\mathrm{plim}\, a_n' b_n = a' b$
  3. $\mathrm{plim}\, C_n^{-1} = C^{-1}$ o ile C^{-1} istnieje

Wnioski z Twierdzenia Słuckiego

Zbieżność według rozkładu

Mówimy, że ciąg zmiennych losowych X_n jest zbieżny według rozkładu do rozkładu prawdopodobieństwa o dystrybuancie $F(X)$, jeżeli

$$\lim_{n\to\infty} F_n(X) = F(X)$$

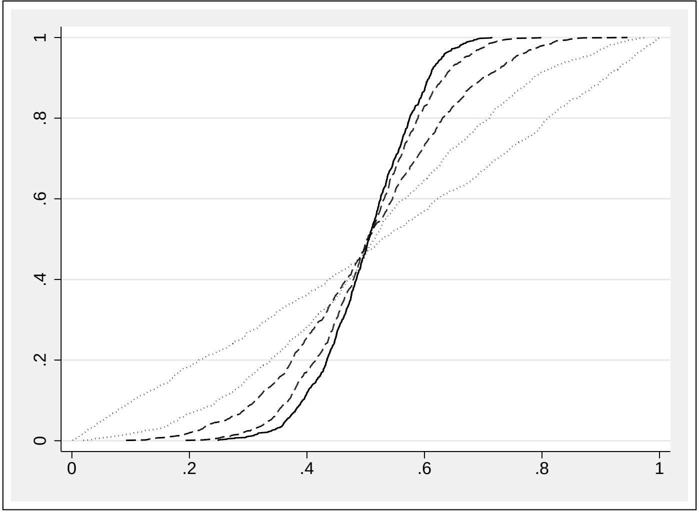
Zbieżność według rozkładu oznaczamy

$$X_n \xrightarrow{D} X$$

- Dodatkowo jeśli $x_n \xrightarrow{D} \eta$, $\operatorname{plim} y_n = y$, oraz x_n i y_n są niezależne to

$$x_n + y_n \xrightarrow{D} \eta + y, \; x_n y_n \xrightarrow{D} \eta' y$$

Przykład. Zbieżność według rozkładu



Zgodność estymatora wektora parametrów MNK

$$\text{plim}(b) = \text{plim}[(X'X)^{-1}X'y] = \text{plim}[(X'X)^{-1}X'(X\beta + \varepsilon)] =$$

- Z własności zbieżności według prawdopodobieństwa

$$= \beta + \text{plim}\left[(X'X)^{-1}X'\varepsilon\right] = \beta + \text{plim}\left[N(X'X)^{-1}\frac{1}{N}X'\varepsilon\right]$$

$$= \beta + \text{plim}\left[\left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{N}\right] = \beta + \text{plim}\left(\frac{X'X}{N}\right)^{-1}\text{plim}\left(\frac{X'\varepsilon}{N}\right)$$

- Z własności 3 i 4 oraz PWL wiemy, że

$$= \beta + E(x_i'x_i)^{-1}E(\underbrace{x_i'\varepsilon_i}_{0}) = \beta$$

Zgodność estymatora wektora parametrów MNK

- Ale

$$cov(x_i, \varepsilon_i) = E(x_i, \varepsilon_i) - E(x_i)\underbrace{E(\varepsilon_i)}_{0} = E(x_i, \varepsilon_i)$$

- Wobec tego do uzyskania zgodności estymatora niezbędne jest by $cov(x_i, \varepsilon_i) = 0$
- To założenie jest słabsze od warunkowej niezależności składnika losowego od macierzy obserwacji
- Chociaż przy spełnionym założeniu o wielowymiarowej normalności składnika losowego są one równoważne

Zgodność estymatora wektora parametrów MNK

- Aby pokazać zgodność estymatora MNK, należy udowodnić że średnie wartości z próby zbiegają do wartości oczekiwanych
- Pokażemy zbieżność dla pojedynczych elementów wektorów $\sum x_i' x_i$ oraz $\sum x_i' \varepsilon_i$
- Oznaczmy je odpowiednio $\sum x_{ri}' x_{si}$ oraz $\sum x_{ri}' \varepsilon_i$

Dowód zgodności (szkic)

- Z PWL wynika, że założenia 3 i 4 są spełnione jeżeli
  1. $var(x_{ri}x_{si}) = \sigma^2_{x_r x_s} < \infty \quad \forall r, s, \; i = 1 \dots n$
  2. $cov(x_{ri}x_{si}, x_{rj}x_{sj}) = 0 \quad \forall i, j = 1 \dots n \wedge i \neq j$

  i analogicznie
  3. $var(x_{ri}\varepsilon_i) = \sigma^2_{x_r \varepsilon} < \infty \quad \forall r, \; i = 1 \dots n$
  4. $cov(x_{ri}\varepsilon_i, x_{rj}\varepsilon_j) = 0 \quad \forall i, j = 1 \dots n \wedge i \neq j$

Zgodność estymatora wektora parametrów MNK - przypadek ogólny

- Warunki zgodności dla estymatora MNK przy heteroscedastyczności i/lub autokorelacji składnika losowego
  1. $\lim_{n\to\infty} \frac{1}{N} var(x_{ri}x_{si}) = \sigma^2_{x_r x_s} < \infty \quad \forall r, s,\ i = 1 \dots n$
  2. $\lim_{n\to\infty} \frac{1}{N} cov(x_{ri}x_{si}, x_{rj}x_{sj}) = 0 \quad \forall i, j = 1 \dots n \wedge i \neq j$
  3. $\lim_{n\to\infty} \frac{1}{N} var(x_{ri}\varepsilon_i) = \sigma^2_{x_r \varepsilon} < \infty \quad \forall r,\ i = 1 \dots n$
  4. $\lim_{n\to\infty} \frac{1}{N} cov(x_{ri}\varepsilon_i, x_{rj}\varepsilon_j) = 0 \quad \forall i, j = 1 \dots n \wedge i \neq j$

Asymptotyczny rozkład estymatora MNK

- Założenie

$$\frac{1}{\sqrt{N}}X'\varepsilon = \sqrt{N}\left(\frac{1}{N}\sum x_i\varepsilon_i\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2 E(x_i'x_i)\right)$$

- więc

$$\sqrt{N}(b-\beta) = \left(\frac{1}{N}X'X\right)^{-1}\left(\frac{1}{N}\sum x_i\varepsilon_i\right) \xrightarrow{D} [E(x_i'x_i)]^{-1}\eta$$

- Z własności rozkładu normalnego i twierdzenia Sluckiego

$$\sqrt{N}(b-\beta) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2 E(x_i'x_i)\right)$$

- Wobec tego

$$b \overset{a}{\sim} \mathcal{N}\left(\beta, \sigma^2(X'X)^{-1}\right)$$

Równoczesność

- O występowaniu równoczesności mówimy, gdy

$$cov(x_i, \varepsilon_i) = E(x_i, \varepsilon_i) \neq 0$$

- Są dwa powody występowania równoczesności
  1. zmienne pominięte
  2. sprzężenie zwrotne
- Konsekwencją występowania równoczesności jest brak zgodności estymatora

Równoczesność

- Mamy model

$$ln(zarobki) = \beta_0 + \beta_1 plec + \beta_2 dosw + \beta_3 dosw2 + \varepsilon$$

- Jeśli pominiemy zmienną $dosw2$ i oszacujemy model

$$ln(zarobki) = \beta_0 + \beta_1 plec + \beta_2 dosw + \eta$$

- To wtedy

$$\eta = \beta_3 dosw2 + \varepsilon$$

W konsekwencji $cov(\eta, dosw) \neq 0$