Metoda największej wiarogodności

- Metoda Największej Wiarogodności (MNW) jest bardziej uniwersalną niż MNK metodą szacowania wartości nieznanych parametrów

- Metoda Największej Wiarogodności (MNW) jest bardziej uniwersalną niż MNK metodą szacowania wartości nieznanych parametrów
- Wymaga spełnienia założeń, które są mniej restrykcyjne niż założenia MNK

- Metoda Największej Wiarogodności (MNW) jest bardziej uniwersalną niż MNK metodą szacowania wartości nieznanych parametrów
- Wymaga spełnienia założeń, które są mniej restrykcyjne niż założenia MNK
- Dzięki temu MNW może być wykorzystywana do estymacji szerszej klasy modeli ekonometrycznych

Wiarogodność

Wiarogodność jest to funkcja $\Theta \rightarrow \mathbb{R}$:

$$L(\theta) = f(\theta, x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n)$$

Wiarogodność

Wiarogodność jest to funkcja $\Theta \longrightarrow \mathbb{R}$:

$$L(\theta) = f(\theta, x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n)$$

- W praktyce będziemy dysponowali próbą n obserwacji. Kluczowym założeniem jest znajomość postaci funkcji $f(\cdot)$

Estymator Metody Największej Wiarogodności

Estymatorem Metody Największej Wiarogodności (MNW) dla wektora nieznanych parametrów θ, szacowanym na podstawie wartości zmiennej zależnej y i wartości zmiennych niezależnych X, nazywamy

$$\hat{\theta} = \underset{\theta \in \Theta}{argmax}\ f_{\theta}(y, X)$$

Estymator Metody Największej Wiarogodności

Estymatorem Metody Największej Wiarogodności (MNW) dla wektora nieznanych parametrów θ, szacowanym na podstawie wartości zmiennej zależnej y i wartości zmiennych niezależnych X, nazywamy

$$\hat{\theta} = \underset{\theta \in \Theta}{argmax} \ f_{\theta}(y, X)$$

- $f(\cdot)$ nazywamy łączną funkcją gęstości prawdopodobieństwa lub funkcją wiarogodności

Estymator Metody Największej Wiarogodności

Estymatorem Metody Największej Wiarogodności (MNW) dla wektora nieznanych parametrów θ, szacowanym na podstawie wartości zmiennej zależnej y i wartości zmiennych niezależnych X, nazywamy

$$\hat{\theta} = \underset{\theta \in \Theta}{argmax}\ f_\theta(y, X)$$

- $f(\cdot)$ nazywamy łączną funkcją gęstości prawdopodobieństwa lub funkcją wiarogodności
- Metoda polega na wybraniu takiej wartości parametru θ, który maksymalizuje $f(\cdot)$ dla zaobserwowanych wartości y oraz X

1. Niezależność obserwacji (losowość próby)

1. Niezależność obserwacji (losowość próby)
2. Identyfikowalność parametrów

1. Niezależność obserwacji (losowość próby)
2. Identyfikowalność parametrów
3. Słaba egzogeniczność

1. Niezależność obserwacji (losowość próby)
2. Identyfikowalność parametrów
3. Słaba egzogeniczność
4. Znajomość postaci funkcji wiarodogności

Niezależność obserwacji

- W przypadku danych przekrojowych jest równoważna losowości próby

Niezależność obserwacji

- W przypadku danych przekrojowych jest równoważna losowości próby
- Jeżeli próba przekrojowa nie jest losowa wymagane są dodatkowe założenia

Niezależność obserwacji

- W przypadku danych przekrojowych jest równoważna losowości próby
- Jeżeli próba przekrojowa nie jest losowa wymagane są dodatkowe założenia
- W przypadku danych czasowych zazwyczaj nie da się utrzymać założenia o losowości próby

Niezależność obserwacji

- W przypadku danych przekrojowych jest równoważna losowości próby
- Jeżeli próba przekrojowa nie jest losowa wymagane są dodatkowe założenia
- W przypadku danych czasowych zazwyczaj nie da się utrzymać założenia o losowości próby
- Jednak, w praktyce, założenie o niezależności komplikuje jedynie dowody własności estymatorów

Niezależność obserwacji

- Przy spełnionym założeniu o niezależności funkcję wiarogodności można zapisać jako

$$L_\theta(y|X) = f_\theta(y|X) = f_\theta(y|X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(y_i|X_i)$$

Niezależność obserwacji

- Przy spełnionym założeniu o niezależności funkcję wiarogodności można zapisać jako

$$L_\theta(y|X) = f_\theta(y|X) = f_\theta(y|X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(y_i|X_i)$$

- Zatem łączna funkcja gęstości jest iloczynem funkcji gęstości dla pojedynczych obserwacji

Identyfikowalność parametrów

- Każdy z parametrów modelu powinien w odmienny sposób wpływać na prawdopodobieństwo zaobserwowania próby

Identyfikowalność parametrów

- Każdy z parametrów modelu powinien w odmienny sposób wpływać na prawdopodobieństwo zaobserwowania próby
- Dzięki tej własności można uzyskać oddzielne oszacowanie dla każdego nieznanego parametru

Identyfikowalność parametrów

- Każdy z parametrów modelu powinien w odmienny sposób wpływać na prawdopodobieństwo zaobserwowania próby
- Dzięki tej własności można uzyskać oddzielne oszacowanie dla każdego nieznanego parametru

Identyfikowalność

Niech θ_0 będzie prawdziwą wielkością wektora parametrów θ. Wówczas istnieje taki zbiór danych, że

$$\exists\,(y,X)\;\forall\theta \neq \theta_0 :\; L_{\theta_0}(X,y) \neq L_{\theta}(X,y)$$

Identyfikowalność parametrów

- W otoczeniu punktu θ_0 funkcja wiarogodności nie może być lokalnie płaska

Identyfikowalność parametrów

- W otoczeniu punktu θ_0 funkcja wiarogodności nie może być lokalnie płaska
- Prawdopodobieństwo zaobserwowania próby (X,y) dla wektora parametrów θ_0 i dla wektora parametrów θ powinno być różne

Identyfikowalność parametrów

- W otoczeniu punktu θ_0 funkcja wiarogodności nie może być lokalnie płaska
- Prawdopodobieństwo zaobserwowania próby (X,y) dla wektora parametrów θ_0 i dla wektora parametrów θ powinno być różne
- Inaczej mówiąc maksimum funkcji wiarogodności powinno być jednoznacznie określone

Identyfikowalność parametrów - przykład

- Przypuśćmy, że obserwujemy zmienną losową

$$y_i = \begin{cases} y_i = 0 & y_i^\star < 0 \\ y_i = 1 & y_i^\star > 0 \end{cases} \text{,gdzie } y_i^\star \sim \mathcal{N}(\mu, \sigma^2)$$

Identyfikowalność parametrów - przykład

- Przypuśćmy, że obserwujemy zmienną losową

$$y_i = \begin{cases} y_i = 0 & y_i^\star < 0 \\ y_i = 1 & y_i^\star > 0 \end{cases} \text{,gdzie } y_i^\star \sim \mathcal{N}(\mu, \sigma^2)$$

- Zmienną nieobserwowaną $y_i^\star$ nazywamy zmienną ukrytą

Identyfikowalność parametrów - przykład

- Przypuśćmy, że obserwujemy zmienną losową

$$y_i = \begin{cases} y_i = 0 & y_i^\star < 0 \\ y_i = 1 & y_i^\star > 0 \end{cases} \text{,gdzie } y_i^\star \sim \mathcal{N}(\mu, \sigma^2)$$

- Zmienną nieobserwowaną $y_i^\star$ nazywamy zmienną ukrytą
- Celem jest oszacowanie wartości parametrów rozkładu zmiennej ukrytej

Identyfikowalność parametrów - przykład

- Przypuśćmy, że obserwujemy zmienną losową

$$y_i = \begin{cases} y_i = 0 & y_i^\star < 0 \\ y_i = 1 & y_i^\star > 0 \end{cases} \text{,gdzie } y_i^\star \sim \mathcal{N}(\mu, \sigma^2)$$

- Zmienną nieobserwowaną $y_i^\star$ nazywamy zmienną ukrytą
- Celem jest oszacowanie wartości parametrów rozkładu zmiennej ukrytej
- Wiemy, że

$$\Pr(y_i = 0) = \Pr(y_i^\star < 0) = \Phi\left(\frac{\mu}{\sigma}\right)$$

$$\Pr(y_i = 1) = \Pr(y_i^\star > 0) = 1 - \Phi\left(\frac{\mu}{\sigma}\right)$$

Identyfikowalność parametrów - przykład

- Niech w próbie jest n_0 obserwacji dla których $y_i = 0$, oraz n_1 obserwacji dla których $y_i = 1$

Identyfikowalność parametrów - przykład

- Niech w próbie jest n_0 obserwacji dla których $y_i = 0$, oraz n_1 obserwacji dla których $y_i = 1$
- Wówczas funkcja wiarogodności ma postać

$$L(\mu_0, \sigma_0) = \left[\Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]^{n_0}\left[1 - \Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]^{n_1}$$

Identyfikowalność parametrów - przykład

- Niech w próbie jest n_0 obserwacji dla których $y_i = 0$, oraz n_1 obserwacji dla których $y_i = 1$
- Wówczas funkcja wiarogodności ma postać

$$L(\mu_0, \sigma_0) = \left[\Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]^{n_0}\left[1 - \Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]^{n_1}$$

- Ale połóżmy $\mu^\star = \alpha\mu_0$ oraz $\sigma^\star = \alpha\sigma_0$ dla dowolnego $\alpha > 0$. Wówczas

$$L(\mu^\star, \sigma^\star) = \left[\Phi\left(\frac{\alpha\mu_0}{\alpha\sigma_0}\right)\right]^{n_0}\left[1 - \Phi\left(\frac{\alpha\mu_0}{\alpha\sigma_0}\right)\right]^{n_1} = L(\mu, \sigma)$$

Identyfikowalność parametrów - przykład

- Niech w próbie jest n_0 obserwacji dla których $y_i = 0$, oraz n_1 obserwacji dla których $y_i = 1$
- Wówczas funkcja wiarogodności ma postać

$$L(\mu_0, \sigma_0) = \left[\Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]^{n_0}\left[1 - \Phi\left(\frac{\mu_0}{\sigma_0}\right)\right]^{n_1}$$

- Ale połóżmy $\mu^\star = \alpha\mu_0$ oraz $\sigma^\star = \alpha\sigma_0$ dla dowolnego $\alpha > 0$. Wówczas

$$L(\mu^\star, \sigma^\star) = \left[\Phi\left(\frac{\alpha\mu_0}{\alpha\sigma_0}\right)\right]^{n_0}\left[1 - \Phi\left(\frac{\alpha\mu_0}{\alpha\sigma_0}\right)\right]^{n_1} = L(\mu, \sigma)$$

- Zatem funkcja nie posiada jednoznacznie wyznaczonego maksimum

Słaba egzogeniczność

- Słaba egzogeniczność w MNK utożsamialiśmy z nieskorelowaniem obserwacji z równoczesnym błędem losowym

Słaba egzogeniczność

- Słaba egzogeniczność w MNK utożsamialiśmy z nieskorelowaniem obserwacji z równoczesnym błędem losowym
- Załóżmy, że wektor parametrów θ można separować na dwie części $\theta = (\Phi, \Psi)$

Słaba egzogeniczność

- Słaba egzogeniczność w MNK utożsamialiśmy z nieskorelowaniem obserwacji z równoczesnym błędem losowym
- Założmy, że wektor parametrów θ można separować na dwie części $\theta = (\Phi, \Psi)$
- Założmy, że te części są niepowiązane, czyli $\phi \in \Phi$, $\psi \in \Psi$ to $\theta \in \Phi \times \Psi$

Słaba egzogeniczność

- Słaba egzogeniczność w MNK utożsamialiśmy z nieskorelowaniem obserwacji z równoczesnym błędem losowym
- Załóżmy, że wektor parametrów θ można separować na dwie części $\theta = (\Phi, \Psi)$
- Załóżmy, że te części są niepowiązane, czyli $\phi \in \Phi$, $\psi \in \Psi$ to $\theta \in \Phi \times \Psi$

Słaba egzogeniczność

Mówimy, że X jest słabo egzogeniczne względem wektora parametrów Ψ jeżeli łączną funkcje gęstości można następująco dekomponować

$$f_\theta(y, X) = f_\Phi(y, X) f_\Psi(X)$$

Słaba egzogeniczność

- Funkcja $f_{\Psi}(X)$ opisuje prawdopodobieństwo zaobserwowania wartości zmiennych egzogenicznych i nie ma wpływu na prawdopodobieństwo zaobserwowania konkretnych wartości parametrów modelu

Słaba egzogeniczność

- Funkcja $f_\Psi(X)$ opisuje prawdopodobieństwo zaobserwowania wartości zmiennych egzogenicznych i nie ma wpływu na prawdopodobieństwo zaobserwowania konkretnych wartości parametrów modelu
- Zatem

$$\max_\theta f_\theta(y, X) = \max_\Phi (y, X) + \max_\Psi (X)$$

Słaba egzogeniczność

- Funkcja $f_\Psi(X)$ opisuje prawdopodobieństwo zaobserwowania wartości zmiennych egzogenicznych i nie ma wpływu na prawdopodobieństwo zaobserwowania konkretnych wartości parametrów modelu
- Zatem
$$\max_\theta f_\theta(y, X) = \max_\Phi (y, X) + \max_\Psi (X)$$
- Więc estymator MNW można obliczyć jako sumę dwóch oszacowań

Słaba egzogeniczność

- Funkcja $f_\Psi(X)$ opisuje prawdopodobieństwo zaobserwowania wartości zmiennych egzogenicznych i nie ma wpływu na prawdopodobieństwo zaobserwowania konkretnych wartości parametrów modelu
- Zatem
$$\max_\theta f_\theta(y, X) = \max_\Phi (y, X) + \max_\Psi (X)$$
- Więc estymator MNW można obliczyć jako sumę dwóch oszacowań
- Wniosek: Wartość parametru ϕ można oszacować na podstawie warunkowej funkcji gęstości $f_\Phi(y|X)$

- W praktyce zamiast funkcji wiarogodności $L(\theta)$ posługujemy się logarytmem funkcji wiarogodnoćci $\ell(\theta)$

- W praktyce zamiast funkcji wiarogodności $L(\theta)$ posługujemy się logarytmem funkcji wiarogodnoćci $\ell(\theta)$
- Jest to uprawnione, ponieważ nie interesuje nas wartość funkcji, a wielkość oszacowań nieznanych parametrów

- W praktyce zamiast funkcji wiarogodności $L(\theta)$ posługujemy się logarytmem funkcji wiarogodnoćci $\ell(\theta)$
- Jest to uprawnione, ponieważ nie interesuje nas wartość funkcji, a wielkość oszacowań nieznanych parametrów
- Logarytm jako monotoniczna transformacja zachowuje ekstrema

- W praktyce zamiast funkcji wiarogodności $L(\theta)$ posługujemy się logarytmem funkcji wiarogodnoćci $\ell(\theta)$
- Jest to uprawnione, ponieważ nie interesuje nas wartość funkcji, a wielkość oszacowań nieznanych parametrów
- Logarytm jako monotoniczna transformacja zachowuje ekstrema
- Jeżeli obserwacje są niezależne to logarytm funkcji wiarogodności dany jest przez

$$\ell_\theta(y|X) = \ln \prod f_\theta(y_i|x_i) = \sum_{i=1}^{N} \ln f_\theta(y_i|x_i) = \sum_{i=1}^{N} \ln \ell(y|X_i\theta)$$

- W praktyce zamiast funkcji wiarogodności $L(\theta)$ posługujemy się logarytmem funkcji wiarogodnoćci $\ell(\theta)$
- Jest to uprawnione, ponieważ nie interesuje nas wartość funkcji, a wielkość oszacowań nieznanych parametrów
- Logarytm jako monotoniczna transformacja zachowuje ekstrema
- Jeżeli obserwacje są niezależne to logarytm funkcji wiarogodności dany jest przez

$$\ell_\theta(y|X) = \ln \prod f_\theta(y_i|x_i) = \sum_{i=1}^{N} \ln f_\theta(y_i|x_i) = \sum_{i=1}^{N} \ln \ell(y|X_i\theta)$$

- Logarytm funkcji gęstości jest sumą logarytmów warunkowych gęstości dla poszczególnych obserwacji

- Estymatory MNW wyznacza się maksymalizując funkcję wiarogodności korzystając z warunków pierwszego i drugiego rzędu na istnienie ekstremum funkcji

- Estymatory MNW wyznacza się maksymalizując funkcję wiarogodności korzystając z warunków pierwszego i drugiego rzędu na istnienie ekstremum funkcji
- Pewien problem może stanowić brak rozwiązań analitycznych

- Estymatory MNW wyznacza się maksymalizując funkcję wiarogodności korzystając z warunków pierwszego i drugiego rzędu na istnienie ekstremum funkcji
- Pewien problem może stanowić brak rozwiązań analitycznych
- Dla programów komputerowych nie stanowi to problemu, gdyż one znajdują pochodne w sposób numeryczny

Przykład

- Znajdź estymatory MNW dla parametrów KMRL

$$y_i = X_i\beta + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Przykład

- Znajdź estymatory MNW dla parametrów KMRL

$$y_i = X_i\beta + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Z założeń modelu wynika, że

$$y_i | x_i \sim \mathcal{N}(x_i\beta, \sigma^2)$$

Przykład

- Znajdź estymatory MNW dla parametrów KMRL

$$y_i = X_i\beta + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Z założeń modelu wynika, że

$$y_i|x_i \sim \mathcal{N}(x_i\beta, \sigma^2)$$

- Więc funkcja gęstości dla pojedynczej obserwacji jest dana przez

$$f_{\beta,\sigma^2}(y_i|x_i) = \frac{1}{\sqrt{2\Pi\sigma^2}} \exp\left(-\frac{(y - X_i\beta)^2}{2\sigma^2}\right)$$

Przykład

- Znajdź estymatory MNW dla parametrów KMRL

$$y_i = X_i\beta + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Z założeń modelu wynika, że

$$y_i | x_i \sim \mathcal{N}(x_i\beta, \sigma^2)$$

- Więc funkcja gęstości dla pojedynczej obserwacji jest dana przez

$$f_{\beta,\sigma^2}(y_i|x_i) = \frac{1}{\sqrt{2\Pi\sigma^2}} \exp\left(- \frac{(y - X_i\beta)^2}{2\sigma^2} \right)$$

- Zatem funkcja wiarogodności ma postać

$$L(\beta, \sigma^2) = \left[\frac{1}{\sqrt{2\Pi\sigma^2}}\right]^n \exp\left(- \frac{1}{2\sigma^2}(y - X_i\beta)'(y - X_i\beta)\right)$$

Przykład - cd

- A logarytm funkcji wiarogodności

$$\ell(\beta, \sigma^2) = -\frac{n}{2}\ln(2\Pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X_i\beta)'(y - X_i\beta)$$

Przykład - cd

- A logarytm funkcji wiarogodności

$$\ell(\beta,\sigma^2) = -\frac{n}{2}\ln(2\Pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X_i\beta)'(y - X_i\beta)$$

- Zatem warunki pierwszego rzędu dane są przez

$$\frac{\partial\ell(\beta,\sigma^2)}{\partial\beta} = -\frac{1}{2\sigma^2}(2X'X\beta - 2X'y) = 0$$

$$\frac{\partial\ell(\beta,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2\sigma^4}(y - X_i\beta)'(y - X_i\beta) = 0$$

Przykład - cd

- A logarytm funkcji wiarogodności

$$\ell(\beta,\sigma^2) = -\frac{n}{2}\ln(2\Pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X_i\beta)'(y - X_i\beta)$$

- Zatem warunki pierwszego rzędu dane są przez

$$\frac{\partial\ell(\beta,\sigma^2)}{\partial\beta} = -\frac{1}{2\sigma^2}(2X'X\beta - 2X'y) = 0$$

$$\frac{\partial\ell(\beta,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2\sigma^4}(y - X_i\beta)'(y - X_i\beta) = 0$$

- Uwaga: warunki drugiego rzędu wyprowadzimy później

Przykład - cd

- A logarytm funkcji wiarogodności

$$\ell(\beta,\sigma^2) = -\frac{n}{2}\ln(2\Pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X_i\beta)'(y - X_i\beta)$$

- Zatem warunki pierwszego rzędu dane są przez

$$\frac{\partial\ell(\beta,\sigma^2)}{\partial\beta} = -\frac{1}{2\sigma^2}(2X'X\beta - 2X'y) = 0$$

$$\frac{\partial\ell(\beta,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2\sigma^4}(y - X_i\beta)'(y - X_i\beta) = 0$$

- Uwaga: warunki drugiego rzędu wyprowadzimy później
- Wynika z tego, że

$$\tilde{\beta} = (X'X)^{-1}X'y$$

$$\tilde{\sigma^2} = \frac{1}{n}(y - X_i\beta)'(y - X_i\beta) = \frac{e'e}{n}$$

Przykład - cd

- WNIOSEK: estymator MNW dla parametru β ma taką samą postać jak estymator MNK; estymatory dla wariancji różnią się

Przykład - cd

- WNIOSEK: estymator MNW dla parametru β ma taką samą postać jak estymator MNK; estymatory dla wariancji różnią się
- Obliczmy wartość oczekiwaną estymatora MNW dla wariancji

$$E(\tilde{\sigma}^2) = E(\frac{e'e}{n}) = \frac{n-k}{n}E(s^2) = \frac{n-k}{n}\sigma^2 \overset{n\to\infty}{\longrightarrow} \sigma^2$$

Przykład - cd

- WNIOSEK: estymator MNW dla parametru β ma taką samą postać jak estymator MNK; estymatory dla wariancji różnią się
- Obliczmy wartość oczekiwaną estymatora MNW dla wariancji

$$E(\tilde{\sigma}^2) = E(\frac{e'e}{n}) = \frac{n-k}{n} E(s^2) = \frac{n-k}{n}\sigma^2 \overset{n\to\infty}{\longrightarrow} \sigma^2$$

- Zatem estymator MWN dla wariancji jest zgodny, ale obciążony w małych próbach

1. Zgodność

1. Zgodność
2. Asymptotyczna normalność

1. Zgodność
2. Asymptotyczna normalność
3. Asymptotyczna efektywność

Zgodność

Zgodność

Estymator nazywamy zgodnym, gdy

$$\forall \delta > 0 \lim_{n \to \infty} \Pr(|\hat{\theta} - \theta| < \delta) = 1$$

Zgodność

Zgodność

Estymator nazywamy zgodnym, gdy

$$\forall\delta > 0 \lim_{n\to\infty} \Pr(|\hat{\theta} - \theta| < \delta) = 1$$

- Inaczej mówiąc estymator dąży według prawdopodobieństwa do prawdziwej wartości parametru, co zapisujemy

$$\hat{\theta} \xrightarrow{p} \theta \qquad\qquad \operatorname{plim} \hat{\theta} = \theta$$

Zgodność

Zgodność

Estymator nazywamy zgodnym, gdy

$$\forall\delta > 0 \lim_{n\to\infty} \Pr(|\hat{\theta} - \theta| < \delta) = 1$$

- Inaczej mówiąc estymator dąży według prawdopodobieństwa do prawdziwej wartości parametru, co zapisujemy

$$\hat{\theta} \xrightarrow{p} \theta \qquad\qquad \operatorname{plim} \hat{\theta} = \theta$$

- Estymator nazywamy zgodnym, gdy wartość oszacowania parametru zbiega do prawdziwej wartości estymatora

Zgodność

Zgodność

Estymator nazywamy zgodnym, gdy

$$\forall \delta > 0 \lim_{n \to \infty} \Pr(|\hat{\theta} - \theta| < \delta) = 1$$

- Inaczej mówiąc estymator dąży według prawdopodobieństwa do prawdziwej wartości parametru, co zapisujemy

$$\hat{\theta} \xrightarrow{p} \theta \qquad \operatorname{plim} \hat{\theta} = \theta$$

- Estymator nazywamy zgodnym, gdy wartość oszacowania parametru zbiega do prawdziwej wartości estymatora
- Bardzie potocznym językiem: jeżeli liczebność próby rośnie to różnica między wartością estymatora a prawdziwą wartością parametru jest nieskończenie bliska 0

Asymptotyczna normalność

- Graniczny (asymptotyczny) rozkład estymatora to rozkład normalny

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0, I^{-1}(\theta))$$

gdzie

$$i(\theta) = \lim_{N \to \infty} \left[\frac{1}{N} \mathbb{I}(\theta) \right] \text{ oraz } \mathbb{I}(\theta) = \mathrm{var} \left[\frac{\partial \ell(\theta)}{\partial \theta} \right] = -E \left[\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right]$$

jest macierzą informacyjną Fishera

Asymptotyczna normalność

- Macierz informacyjna jest równa wariancji gradientu logarytmu funkcji wiarogodności

Asymptotyczna normalność

- Macierz informacyjna jest równa wariancji gradientu logarytmu funkcji wiarogodności
- lub minus wartości oczekiwanej jej Hessianu

Asymptotyczna normalność - przykład

- Znajdziemy gradient i Hessian dla KMRL

Asymptotyczna normalność - przykład

- Znajdziemy gradient i Hessian dla KMRL
- Gradient jest równy

$$\frac{\partial \ell(\beta, \sigma^2)}{\partial \beta} = -\frac{1}{2\sigma^2}(2X'X\beta - 2X'y) = \frac{1}{\sigma^2}X'\varepsilon$$

$$\frac{\partial \ell(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2\sigma^4}(y - X_i\beta)'(y - X_i\beta) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^2}\varepsilon'\varepsilon$$

Asymptotyczna normalność - przykład

- Wariancja wektora pochodnych względem β wynosi

$$\mathrm{var}\left(\frac{\partial \ell(\beta, \sigma^2)}{\partial \beta}\right) = \frac{1}{\sigma^2} X' \mathrm{var}(\varepsilon) X = \frac{1}{\sigma^2} X' X$$

Asymptotyczna normalność - przykład

- Wariancja wektora pochodnych względem β wynosi

$$\operatorname{var}\left(\frac{\partial\ell(\beta,\sigma^2)}{\partial\beta}\right)=\frac{1}{\sigma^2}X'\operatorname{var}(\varepsilon)X=\frac{1}{\sigma^2}X'X$$

- Wariancja wektora pochodnych względem σ^2 wynosi

$$\operatorname{var}\left(\frac{\partial\ell(\beta,\sigma^2)}{\partial\sigma^2}\right)=\operatorname{var}\left(-\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right)=\frac{1}{4\sigma^8}N2\sigma^4=\frac{N}{2\sigma^4}$$

Asymptotyczna normalność - przykład

- Wariancja wektora pochodnych względem β wynosi

$$\mathrm{var}\left(\frac{\partial \ell(\beta,\sigma^2)}{\partial \beta}\right) = \frac{1}{\sigma^2} X' \mathrm{var}(\varepsilon) X = \frac{1}{\sigma^2} X' X$$

- Wariancja wektora pochodnych względem σ^2 wynosi

$$\mathrm{var}\left(\frac{\partial \ell(\beta,\sigma^2)}{\partial \sigma^2}\right) = \mathrm{var}\left(-\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^2}\varepsilon'\varepsilon\right) = \frac{1}{4\sigma^8} N 2\sigma^4 = \frac{N}{2\sigma^4}$$

- Ponieważ $E\left(\frac{1}{\sigma^2} X'\varepsilon\right) = 0$, więc kowariancja między pochodnymi wynosi zero

Asymptotyczna normalność - przykład

- Zatem Hessian jest równy

$$\mathbb{I}(\beta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

Asymptotyczna normalność - przykład

- Te same wyniki uzyskamy obliczając wartości oczekiwane elementów Hessianu

$$\frac{\partial \ell(\beta,\sigma^2)}{\partial\beta\partial\beta'} = -\frac{1}{2\sigma^4}X'X$$

$$\frac{\partial \ell(\beta,\sigma^2)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}X'\varepsilon$$

$$\frac{\partial \ell(\beta,\sigma^2)}{\partial\sigma^2\partial\sigma^2} = \frac{N}{2}\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\varepsilon'\varepsilon$$

Asymptotyczna efektywność

Twierdzenie Rao-Cramera

Jeżeli estymator $\hat{\theta}$ jest zgodny to jego asymptotyczna wariancja jest nie mniejsza niż dolne ograniczenie Rao-Cramera

$$\lim_{N\to\infty} \mathrm{var}\left[\sqrt{n}(\tilde{\theta}-\theta)\right] \geq i^{-1}(\theta) = \lim_{N\to\infty} \frac{1}{N}\mathbb{I}(\theta)$$

Asymptotyczna efektywność

Twierdzenie Rao-Cramera

Jeżeli estymator $\hat{\theta}$ jest zgodny to jego asymptotyczna wariancja jest nie mniejsza niż dolne ograniczenie Rao-Cramera

$$\lim_{N\to\infty} \operatorname{var}\left[\sqrt{n}(\tilde{\theta}-\theta)\right] \geq i^{-1}(\theta) = \lim_{N\to\infty} \frac{1}{N}\mathbb{I}(\theta)$$

- Estymatory MNW są asymptotycznie efektywne ponieważ ich wariancja zbiega do dolnego ograniczenia Rao-Cremera

Asymptotyczna efektywność

Twierdzenie Rao-Cramera

Jeżeli estymator $\hat{\theta}$ jest zgodny to jego asymptotyczna wariancja jest nie mniejsza niż dolne ograniczenie Rao-Cramera

$$\lim_{N\to\infty} \mathrm{var}\left[\sqrt{n}(\tilde{\theta}-\theta)\right] \geq i^{-1}(\theta) = \lim_{N\to\infty} \frac{1}{N}\mathbb{I}(\theta)$$

- Estymatory MNW są asymptotycznie efektywne ponieważ ich wariancja zbiega do dolnego ograniczenia Rao-Cremera
- Zatem estymatory MNW są estymatorami zgodnymi o minimalnej wariancji

- Na podstawie własności estymatorów MNW można wyprowadzić rozkłady statystyk testowych dla ogólnych hipotez

- Na podstawie własności estymatorów MNW można wyprowadzić rozkłady statystyk testowych dla ogólnych hipotez
- Zakładamy, że hipoteza ma postać (nie)liniowego układu równań

$$H_0 : \begin{cases} h_1(\theta) = 0 \\ \vdots \\ h_q(\theta) = 0 \end{cases}$$

- Na podstawie własności estymatorów MNW można wyprowadzić rozkłady statystyk testowych dla ogólnych hipotez
- Zakładamy, że hipoteza ma postać (nie)liniowego układu równań

$$H_0 : \begin{cases} h_1(\theta) = 0 \\ \vdots \\ h_q(\theta) = 0 \end{cases}$$

- Przyjmując $h(\theta) = (h_1(\theta), \ldots, h_q(\theta))$ zapisujemy układ

$$H_0 : \ h(\theta) = 0$$

- Na podstawie własności estymatorów MNW można wyprowadzić rozkłady statystyk testowych dla ogólnych hipotez
- Zakładamy, że hipoteza ma postać (nie)liniowego układu równań

$$H_0: \begin{cases} h_1(\theta) = 0 \\ \vdots \\ h_q(\theta) = 0 \end{cases}$$

- Przyjmując $h(\theta) = (h_1(\theta), \ldots, h_q(\theta))$ zapisujemy układ

$$H_0: \; h(\theta) = 0$$

- Hipotezy należy sformułować w taki sposób, aby macierz pierwszych pochodnych $h'(\theta) = \frac{\partial h(\theta)}{\partial \theta}$ miała pełen rząd

- Jest to w praktyce najłatwiejszy do przeprowadzenia test

- Jest to w praktyce najłatwiejszy do przeprowadzenia test
- Ale wymaga oszacowania dwóch modeli

- Jest to w praktyce najłatwiejszy do przeprowadzenia test
- Ale wymaga oszacowania dwóch modeli
- Weryfikowane jest J ograniczeń

$$LR = 2(\ell(\tilde{\theta}) - \ell(\tilde{\theta}_R)) \xrightarrow{D} \chi^2_J$$

- $\ell(\tilde{\theta})$ jest logarytmem funkcji wiarogodności dla modelu bez ograniczeń
- $\ell(\tilde{\theta}_R)$ jest logarytmem funkcji wiarogodności dla modelu ze spełnionymi ograniczeniami

- Idea testu bazuje na spostrzeżeniu, iż łatwiej jest zmaksymalizować funkcję bez ograniczeń, niż z narzuconymi na parametry restrykcjami

- Idea testu bazuje na spostrzeżeniu, iż łatwiej jest zmaksymalizować funkcję bez ograniczeń, niż z narzuconymi na parametry restrykcjami
- Nazwa testu wywodzi się z faktu, że statystykę testową można zapisać jako

$$LR = 2\ln\left(\frac{L(\tilde{\theta})}{L(\tilde{\theta}_R)}\right)$$

- Idea testu bazuje na spostrzeżeniu, iż łatwiej jest zmaksymalizować funkcję bez ograniczeń, niż z narzuconymi na parametry restrykcjami
- Nazwa testu wywodzi się z faktu, że statystykę testową można zapisać jako

$$LR = 2 \ln \left(\frac{L(\tilde{\theta})}{L(\tilde{\theta_R})} \right)$$

- Wadą testu jest konieczność oszacowania dwóch modeli

- Do obliczenia wartości statystyki Walda wystarczająca jest znajomość oszacowań modelu bez narzuconych ograniczeń

$$W = h'(\theta)\left[H(\theta)\mathbb{I}^{-1}(\theta)H'(\theta)\right]^{-1}h(\theta) \xrightarrow{D} \chi^2_J$$

- Do obliczenia wartości statystyki Walda wystarczająca jest znajomość oszacowań modelu bez narzuconych ograniczeń

$$W = h'(\theta)\left[H(\theta)\mathbb{I}^{-1}(\theta)H'(\theta)\right]^{-1}h(\theta) \xrightarrow{D} \chi^2_J$$

- Macierz w nawiasach kwadratowych jako macierz wariancji jest dodatnio określona

- Do obliczenia wartości statystyki Walda wystarczająca jest znajomość oszacowań modelu bez narzuconych ograniczeń

$$W = h'(\theta)\left[H(\theta)\mathbb{I}^{-1}(\theta)H'(\theta)\right]^{-1}h(\theta) \xrightarrow{D} \chi^2_J$$

- Macierz w nawiasach kwadratowych jako macierz wariancji jest dodatnio określona
- Zatem statystyka $W = 0$, gdy ograniczenia narzucone na parametry są spełnione

- Do obliczenia wartości statystyki Walda wystarczająca jest znajomość oszacowań modelu bez narzuconych ograniczeń

$$W = h'(\theta)\left[H(\theta)\mathbb{I}^{-1}(\theta)H'(\theta)\right]^{-1}h(\theta) \xrightarrow{D} \chi^2_J$$

- Macierz w nawiasach kwadratowych jako macierz wariancji jest dodatnio określona
- Zatem statystyka $W = 0$, gdy ograniczenia narzucone na parametry są spełnione
- Statystyka Walda nie jest niezmiennicza ze względu na sposób zapisania hipotezy zerowej

- Statystyka mnożników Lagrangea wymaga znajomości estymatora dla modelu z narzuconymi ograniczeniami

$$LM = \frac{\partial \ell(\theta)}{\partial \theta}\bigg|_{\theta=\theta_R} \mathbb{I}^{-1}(\theta_R) \frac{\partial \ell(\theta)}{\partial \theta}\bigg|_{\theta=\theta_R} \xrightarrow{D} \chi^2_J$$

- Statystyka mnożników Lagrangea wymaga znajomości estymatora dla modelu z narzuconymi ograniczeniami

$$LM = \left.\frac{\partial \ell(\theta)}{\partial \theta}\right|_{\theta=\theta_R} \mathbb{I}^{-1}(\theta_R) \left.\frac{\partial \ell(\theta)}{\partial \theta}\right|_{\theta=\theta_R} \xrightarrow{D} \chi^2_J$$

- Dla maksimum bez ograniczeń wartość gradientu wynosi zero

- Statystyka mnożników Lagrangea wymaga znajomości estymatora dla modelu z narzuconymi ograniczeniami

$$LM = \frac{\partial \ell(\theta)}{\partial \theta}\bigg|_{\theta=\theta_R} \mathbb{I}^{-1}(\theta_R) \frac{\partial \ell(\theta)}{\partial \theta}\bigg|_{\theta=\theta_R} \xrightarrow{D} \chi^2_J$$

- Dla maksimum bez ograniczeń wartość gradientu wynosi zero
- Zatem wartość gradientu niesie informację o spełnieniu ograniczeń przez parametry modelu

- W modelach szacowanych MNW nie są szacowane sumy kwadratów

- W modelach szacowanych MNW nie są szacowane sumy kwadratów
- Miary dopasowania są przybliżonymi statystykami określanymi jako pseudo-R^2

- W modelach szacowanych MNW nie są szacowane sumy kwadratów
- Miary dopasowania są przybliżonymi statystykami określanymi jako pseudo-R^2
- Najczęściej stosowaną miarą jest R^2 Mc-Faddena

$$R^2_{McFadden} = 1 - \frac{\ell(\theta)}{\ell(\theta_0)}$$

- W modelach szacowanych MNW nie są szacowane sumy kwadratów
- Miary dopasowania są przybliżonymi statystykami określanymi jako pseudo-R^2
- Najczęściej stosowaną miarą jest R^2 Mc-Faddena

$$R^2_{McFadden} = 1 - \frac{\ell(\theta)}{\ell(\theta_0)}$$

- Podobnie jak $adj - R^2$ istnieje

$$\bar{R}^2_{McFadden} = 1 - \frac{\ell(\theta) - K}{\ell(\theta_0)}$$

gdzie L_0 to logarytm funkcji wiarogodności modelu ze stałą jako jedyną zmienną objasniajacą