9. Nieliniowa Metoda Najmniejszych Kwadratów

9.1. Regresja nieliniowa

Modele liniowe są w ekonomii często niewystarczające jako przybliżenie prawdziwych zależności między zmiennymi. Uogólniając model liniowy możemy estymować parametry równania

$$y_i = h(\boldsymbol{x}_i, \boldsymbol{\beta}) + \varepsilon_i$$

gdzie $h(\bullet, \bullet)$ jest pewną funkcją, $\boldsymbol{x}_i$ jest wektorem zmiennych egzogenicznych, $\boldsymbol{\beta}$ wektorem parametrów estymowanaych, zaś ε_i błędem losowym. Taka forma równania regresji, choć ogólniejsza od regresji liniowej nie wyczerpuje jednak wszystkich możliwości, na przykład zależności $y_i = \frac{\boldsymbol{\alpha}}{\boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i}$ nie da zapisać w ten sposób. Podobnie jak w przypadku MNK także w przypadku nieliniowej regresji za estymator parametru $\boldsymbol{\beta}$ przyjmujemy takie $\boldsymbol{b}$, które zapewnia najlepsze dopasowanie $\widehat{y}_i = h(\boldsymbol{x}_i, \boldsymbol{b})$ do y_i. Za kryterium dopasowania przyjmujemy tak jak w przypadku MNK sumę kwadratów reszt.

$$S(\boldsymbol{b}) = \frac{1}{2}\sum_{i=1}^{n}\left[y_i - h(\boldsymbol{x}_i, \boldsymbol{b})\right]^2 = \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 = \frac{1}{2}\sum_{i=1}^{n}\varepsilon_i^2$$

Warunki pierwszego rzędu dla problemu

$$\min_{\boldsymbol{b}} S(\boldsymbol{b})$$

mają postać układu równań nieliniowych:

$$\frac{\partial S(\boldsymbol{b})}{\partial \boldsymbol{b}} = -\sum_{i=1}^{n}\left[y_i - h(\boldsymbol{x}_i, \boldsymbol{b})\right]\frac{\partial h(\boldsymbol{x}_i, \boldsymbol{b})}{\partial \boldsymbol{b}}$$

Układ ten z regóły nie ma analitycznego analitycznego rozwiązania. W związku z tym pojawiają się dwa problemy. Po pierwsze jak przeanalizować własności estymatora $NMNK$ nie znając jego nalitycznej formy a po drugie w jaki sposób można znaleźć wartość tego estymatora.

9.2. Zlinearyzowana regresja

Własności estymatora $NMNK$ zbadamy posługując się zlinearyzowaną wersją nieliniowej regresji. Zamiast rozważać własności nieliniowej funkcji

$$y_i = h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right) + \varepsilon_i$$

skorzystamy z rozwinięcia Taylora, uzyskając liniowej aproksymację tej funkcji. Dla pewnego z góry ustalonego $\boldsymbol{\beta}^0$, mamy

$$\begin{aligned} h\left(\boldsymbol{x}_i, \boldsymbol{\beta}\right) &\approx h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right) + \frac{\partial h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^0}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^0\right) \\ &= h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right) - \frac{\partial h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^0}\boldsymbol{\beta}^0 + \frac{\partial h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^0}\boldsymbol{\beta} \end{aligned}$$

Dla niewielkich wartości $\boldsymbol{\beta} - \boldsymbol{\beta}^0$ taka liniowa funkcja będzie dobrym przybliżeniem prawdziwej wartości $h\left(\boldsymbol{x}_i, \boldsymbol{\beta}\right)$. Oznaczmy jako $\boldsymbol{x}_i^0 = \frac{\partial h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^0}$, a wtedy

$$y_i \approx h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right) - \boldsymbol{x}_i^0\boldsymbol{\beta}^0 + \boldsymbol{x}_i^0\,\boldsymbol{\beta} + \varepsilon_i$$

Jednak tak zdefiniowana funkcja nie ma jeszcze postaci regresji liniowej, ponieważ po prawej stronie znajdują się nie związane z nieznanymi parametrami elementy $h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)$, $\boldsymbol{x}_i^0\boldsymbol{\beta}^0$. Można je jednak przenieść na lewą stronę i definiując $y_i^0 = y_i - h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right) + \boldsymbol{x}_i^0\boldsymbol{\beta}^0$ otrzymujemy

$$y_i^0 \approx \boldsymbol{x}_i^0\,\boldsymbol{\beta} + \varepsilon_i$$

które to równanie ma już postać klasycznego równania regresji. Tak zwanym pseudoregresorami w tym równaniu są elementy wektora $\boldsymbol{x}_i^0 = \frac{\partial h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^0}$, a więc pierwsze pochodne $h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)$ względem elementów wektora $\boldsymbol{\beta}^0$. Regresję tę można zapisać w postaci macierzowej

$$\boldsymbol{y}^0 = \boldsymbol{X}^0\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{9.1}$$

gdzie $\boldsymbol{y}^0 = \begin{bmatrix} y_1^0 \\ \vdots \\ y_n^0 \end{bmatrix}$, $\boldsymbol{X}^0 = \begin{bmatrix} \boldsymbol{x}_1^0 \\ \vdots \\ \boldsymbol{x}_n^0 \end{bmatrix}$.

Odpowiednikiem tej regresji będzie wyestymowane równanie postaci

$$\boldsymbol{y}^0 = \boldsymbol{X}^0\boldsymbol{b} + \boldsymbol{e}.$$

Estymator $NMNK$ minimalizuje sumę kwadratów reszt

$$S(\boldsymbol{b}) = \frac{1}{2}\sum_{i=1}^{n}\left[y_i - h(\boldsymbol{x}_i, \boldsymbol{b})\right]^2 = \frac{1}{2}\sum_{i=1}^{n} e_i^2$$

Warunki pierwszego rzędu dla tego estymatora można zapisać jako

$$\frac{\partial S(\boldsymbol{b})}{\partial \boldsymbol{b}} = -\sum_{i=1}^{n}\left[y_i - h(\boldsymbol{x}_i, \boldsymbol{b})\right]\frac{\partial h(\boldsymbol{x}_i, \boldsymbol{b})}{\partial \boldsymbol{b}} \approx -\sum_{i=1}^{n}\left(\boldsymbol{y}^0 - \boldsymbol{X}^0\boldsymbol{b}\right)\frac{\partial h\left(\boldsymbol{x}_i, \boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^0} = -\boldsymbol{X}^{0\prime}\boldsymbol{e} = \boldsymbol{0}$$

Warunek ortogonalności wynikającego z tego układu równań jest identyczny do tego, który pojawia się w kontekście MNK. Korzystając z (9.1) otrzymujemy układ równań liniowych

$$-\boldsymbol{X}^{0\prime}\boldsymbol{e}^0 = -\boldsymbol{X}^{0\prime}\left(\boldsymbol{y}^0 - \boldsymbol{X}^0\boldsymbol{b}\right) = \boldsymbol{0}$$

który po rozwiązaniu daje nam estymator dla regresji zlinearyzowanej $\boldsymbol{b} = \left(\boldsymbol{X}^{0\prime}\boldsymbol{X}^0\right)^{-1}\boldsymbol{X}^0\boldsymbol{y}^0$. Estymator ten jest *bliski* estymatorowi $NMNK$ parametru $\boldsymbol{\beta}$ pod warunkiem, że $\boldsymbol{y}^0$ i $\boldsymbol{X}^0$ zostały policzone dla $\boldsymbol{\beta}^0$ *bliskiego* $\boldsymbol{\beta}$, który uzyskalibyśmy znajdujując miniumum dla $S(\boldsymbol{\beta})$. Oczywiście w praktyce takie $\boldsymbol{\beta}^0$ trzeba jakoś znaleźć i w związku z tym estymator zlinearyzowany będzie służył nam głównie do rozważań teoretycznych.

9.3. Asymptotyczne własności nieliniowego estymatora metody najmniejszych kwadratów

Warunki zbieżności i normalności wynikają z własności regresji zlinearyzowanej, przy czym wyrażone są poprzez własności macierzy pseudoregresorów $\boldsymbol{X}^0$ policzonych dla prawdziwej wielkości parametru $\boldsymbol{\beta}$. W związku z tym warunki na zbieżność estymatora są identyczne jak analogiczne dla MNK.

Dla zbieżności estymatora konieczne warunkiem jest, by

$$\operatorname{plim}\left(\frac{\boldsymbol{X}^{0\prime}\boldsymbol{X}^0}{n}\right) = \operatorname{plim} n^{-1}\sum_{i=1}^{n}\left(\frac{\partial h\left(\boldsymbol{x}_i,\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}}\right)\left(\frac{\partial h\left(\boldsymbol{x}_i,\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}'}\right) = \boldsymbol{Q}_0 \tag{9.2}$$

gdzie $\boldsymbol{Q}_0$ jest macierzą skończoną oraz asymptotyczny brak skorelowania między $\boldsymbol{X}^0$ i $\boldsymbol{\varepsilon}$:

$$\operatorname{plim}\left(\frac{\boldsymbol{X}^{0\prime}\boldsymbol{\varepsilon}}{n}\right) = \operatorname{plim} n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i^0\varepsilon_i = 0 \tag{9.3}$$

Asymptotyczny rozkład normalny estymator $NMNK$ posiada, gdy dodatkowo

$$\frac{1}{\sqrt{n}}\boldsymbol{X}^{0\prime}\,\boldsymbol{\varepsilon} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{x}_i^0\varepsilon_i \xrightarrow{D} N\left(0,\sigma^2\boldsymbol{Q}_0\right) \tag{9.4}$$

Podobnie jak w przypadku MNK zgodnym estymatorem wariancji błędu losowego będzie

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left[y_i - h\left(\boldsymbol{x}_i,\boldsymbol{b}\right)\right]^2}{n} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n}$$

przy czym jest to estymator zgodny. Zaleca się stosowanie poprawki na ilość stopni swobody, choć w małych próbach estymator ten jest i tak obciążony. W wersji z poprawką estymator ten ma postać

$$\widetilde{\sigma}^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k}$$

Twierdzenie 9.1 *Dla pseudoregresorów spełniający założenia 9.2, 9.3, 9.4 estymator asymptotyczny rozkład estymatora Nieliniowej Metody Najmniejszych Kwadratów ma postać*

$$\sqrt{n}\left(\boldsymbol{b}-\boldsymbol{\beta}\right)\xrightarrow{D} N\left[0,\sigma^2\boldsymbol{Q}_0^{-1}\right],$$

gdzie $\boldsymbol{Q}_0=\operatorname{plim}\left(\frac{\boldsymbol{X}^{0\prime}\boldsymbol{X}^0}{n}\right)$. *Zbieżnym estymatorem asymptotycznej wariancji estymatora* $\boldsymbol{b}$ *jest*

$$\widehat{\boldsymbol{\Sigma}}_{bb}=\widehat{\sigma}^2\left(\boldsymbol{X}^{0\prime}\boldsymbol{X}^0\right)^{-1}$$

Dowód. *Podobnie jak w przypadku* MNK.

$$\begin{aligned}
\sqrt{n}\left(\boldsymbol{b}-\boldsymbol{\beta}\right) &= \sqrt{n}\left[\left(\boldsymbol{X}^{0\prime}\boldsymbol{X}^0\right)^{-1}\boldsymbol{X}^0\boldsymbol{y}^0-\boldsymbol{\beta}\right]\\
&= \sqrt{n}\left[\left(\boldsymbol{X}^{0\prime}\boldsymbol{X}^0\right)^{-1}\boldsymbol{X}^{0\prime}\left(\boldsymbol{X}^0\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)-\boldsymbol{\beta}\right]\\
&= \sqrt{n}\left[\left(\boldsymbol{X}^{0\prime}\boldsymbol{X}^0\right)^{-1}\boldsymbol{X}^{0\prime}\boldsymbol{\varepsilon}\right]\\
&= \left(\frac{\boldsymbol{X}^{0\prime}\boldsymbol{X}^0}{n}\right)^{-1}\frac{1}{\sqrt{n}}\boldsymbol{X}^{0\prime}\boldsymbol{\varepsilon}\xrightarrow{D} N\left[0,\sigma^2\boldsymbol{Q}_0^{-1}\right]
\end{aligned}$$

■

9.4. Testowanie hipotez w modelu nieliniowym

Rozpatrzmy ogólną hipotezę nieliniową postaci

$$H_0:\boldsymbol{H}\left(\boldsymbol{\beta}\right)=\boldsymbol{h},$$

która ma postać układu g równań nieliniowych. Oznaczmy pochodną funkcji wektorowej $\boldsymbol{H}\left(\boldsymbol{\beta}\right)$, jako

$$\boldsymbol{C}\left(\boldsymbol{\beta}\right)=\frac{\partial\boldsymbol{H}\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}'}$$

i zauważmy, że rozwinięcie Taylora $\boldsymbol{H}(\boldsymbol{\beta})$ wokół $\boldsymbol{b}$ ma postać

$$\boldsymbol{H}(\boldsymbol{b}) \approx \boldsymbol{H}(\boldsymbol{\beta}) + \frac{\partial \boldsymbol{H}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}(\boldsymbol{b} - \boldsymbol{\beta})$$

a więc

$$\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{H}(\boldsymbol{\beta}) \approx \frac{\partial \boldsymbol{H}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}(\boldsymbol{b} - \boldsymbol{\beta})$$

i w konsekwencji przy założeniu prawdziwości H_0 otrzymujemy, że

$$\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h} \approx \frac{\partial \boldsymbol{H}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}(\boldsymbol{b} - \boldsymbol{\beta}) = \boldsymbol{C}(\boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta})$$

Wariancja $\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h}$ będzie więc wynosić

$$\mathrm{Var}[\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h}] = \boldsymbol{C}(\boldsymbol{\beta})\,\mathrm{Var}(\boldsymbol{b} - \boldsymbol{\beta})\,\boldsymbol{C}'(\boldsymbol{\beta})$$

Jeśli estymator $\boldsymbol{b}$ jest zbieżny według prawdopodobieństwa i ma asymptotyczny rozkład normalny, to

$$\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h} \xrightarrow{D} N\left(0, \boldsymbol{C}(\boldsymbol{\beta})\,\mathrm{Var}(\boldsymbol{b})\,\boldsymbol{C}'(\boldsymbol{\beta})\right)$$

a

$$[\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h}]'\left[\boldsymbol{C}(\boldsymbol{\beta})\,\mathrm{Var}(\boldsymbol{b})\,\boldsymbol{C}'(\boldsymbol{\beta})\right]^{-1}[\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h}] \xrightarrow{D} \chi_g$$

Zwykle $\boldsymbol{C}(\boldsymbol{\beta})$ i $\mathrm{Var}(\boldsymbol{b})$ są nieznane jednak można je zastąpić w powyższej formie kwadratowej zbieżnymi estymatorami $\boldsymbol{C}(\boldsymbol{b})$ i $\widehat{\Sigma}$, otrzymując ostateczną formę statystyki

$$W = [\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h}]'\left[\boldsymbol{C}(\boldsymbol{b})\,\widehat{\Sigma}\boldsymbol{C}'(\boldsymbol{\beta})\right]^{-1}[\boldsymbol{H}(\boldsymbol{b}) - \boldsymbol{h}] \xrightarrow{D} \chi_g$$

Statystyka o tej formie nazywana jest statystyką Walda.