8.
Identyfikacja liniowych obiektów dynamicznych metodą zmiennych instrumentalnych

1. Koncepcja metody

Gdy w zadaniu identyfikacji liniowego obiektu dynamicznego występuje skorelowanie zakłóceń można zastosować metodę zmiennych instrumentalnych (*ang. IV - instrumental variable method*). Polega ona na zastąpieniu estymatora LS, estymatorem postaci

$$p_N^{(IV)} = (\Psi_N^T \Phi_N)^{-1} \Psi_N^T Y_N \tag{1}$$

gdzie

$$\Psi_N = (\psi_1, \psi_2, ..., \psi_N)^T \tag{2}$$

jest odpowiednią dodatkową macierzą oraz

$$\psi_k = (\psi_{k,1}, ..., \psi_{k,n}, \psi_{k,n+1}, ..., \psi_{k,n+m+1})^T \tag{3}$$

jest wektorem o tym samym wymiarze co wektor φ_k. Macierz Ψ_N jest zatem macierzą o strukturze i wymiarach zgodnych z macierzą Φ_N

$$\dim \Psi_N = \dim \Phi_N \tag{4}$$

Zawiera ona tzw. *zmienne instrumentalne* $\psi_{k,i}$ (instrumenty). Stąd jest nazywana *macierzą zmiennych instrumentalnych*. Procedura identyfikacji systemu w oparciu o metodę zmiennych instrumentalnych wymaga zatem generacji macierzy Ψ_N. Aby ustalić jakie własności powinny spełniać zmienne instrumentalne, po to aby estymator IV dany wzorem (1) był zgodny, mnożymy lewostronnie równanie pomiarów przez Ψ_N^T

$$\Psi_N^T Y_N = \Psi_N^T \Phi_N p^* + \Psi_N^T Z_N \tag{5}$$

Jeżeli wejście $\{u_k\}$ jest procesem iid (co w ogólności będziemy zakładać), przy założeniu asymptotycznej stabilności systemu wyjście $\{y_k\}$ jest procesem ergodycznym. Zachodzi

$$\Psi_N^T Y_N = \Psi_N^T \Phi_N p_N^{(IV)} \tag{6}$$

Wstawiając (6) do (5) po prostych przekształceniach otrzymujemy

$$\left(\frac{1}{N}\Psi_N^T\Phi_N\right)\left(p_N^{(IV)}-p^*\right)=\left(\frac{1}{N}\Psi_N^T Z_N\right)$$

Wynikają stąd dwa podstawowe postulaty jakie równocześnie muszą spełniać zmienne instrumentalne.
Postulat I

Instrumenty winny być takie, że istnieje granica $P\lim_{N\to\infty}(\frac{1}{N}\Psi_N^T\Phi_N)$

$$P\lim_{N\to\infty}(\frac{1}{N}\Psi_N^T\Phi_N)=\mathbf{E}\psi_k\varphi_k^T \tag{7}$$

i jest ona macierzą nieosobliwą

$$\det\{\mathbf{E}\psi_k\varphi_k^T\}\neq 0 \tag{8}$$

(tj. skorelowane z elementami zawartymi w uogólnionym wektorze wejść φ_k, gdy dodatkowo zachodzi jeden z warunków $\mathbf{E}\psi_k=0$ lub $\mathbf{E}\varphi_k=0$; wtedy $cov(\psi_k,\varphi_k)$ jest macierzą niezerową).
Postulat II

Instrumenty powinny być jednocześnie takie, że

$$P\lim_{N\to\infty}(\frac{1}{N}\Psi_N^T Z_N)=\mathbf{E}\psi_k z_k \tag{9}$$

oraz

$$\mathbf{E}\psi_k z_k=0 \tag{10}$$

(tj. nieskorelowane z zakłóceniami, o ile $\mathbf{E}z_k=0$; wtedy $cov(\psi_k,z_k)=0$). Przy zachodzeniu warunku (10) oraz założeniu $\mathbf{E}z_k=0$ zachodzi oczywiście

$$\mathbf{E}\psi_k z_k=\mathbf{E}\psi_k\mathbf{E}z_k=0$$

Uwaga: Dla skuteczności metody zmiennej pomocniczej wystarczyłoby wymaganie, aby macierz graniczna w warunku (7) była dowolną macierzą nieosobliwą, a wektor graniczny w warunku (9) był wektorem zerowym. Świadomie rezygnujemy tutaj z takiej ogólności bowiem w pracy wykazywać będziemy istnienie podanych tu, konkretnych wielkości granicznych, korzystając z ergodyczności odpowiednich procesów.

Przy spełnieniu powyższych postulatów otrzymujemy

$$P\lim(p_N^{(IV)}) = p^*$$

a zatem zgodność estymatora IV niezależnie od budowy korelacyjnej zakłóceń.

Do najbardziej popularnych znanych w literaturze metod generacji macierzy Ψ_N zapewniających spełnienie warunków (7)–(9) należą:

- przypisanie zmiennym pomocniczym wartości wejść w chwilach poprzednich

$$\psi_k = (u_{k-1}, ..., u_{k-n}, u_{k-n-1}, .., u_{k-n-m-1}) \tag{11}$$

- zastosowanie dowolnej filtracji liniowej sygnału wejściowego

$$\psi_{k,j} = \sum_{i=0}^{\infty} h_{j,i} u_{k-i} \tag{12}$$

- iteracyjne wyznaczanie elementów wektora ψ_k związanych z autoregresją w opisie obiektu, tzn. wartości $\psi_{k,1}, ..., \psi_{k,R}$, jako wyjścia modelu obiektu $\widehat{y_k}$ dla dotychczas zidentyfikowanych parametrów

$$\psi_k = (-\widehat{y_{k-1}}, ..., -\widehat{y_{k-R}}, u_k, u_{k-1}, ...u_{k-S}); \tag{13}$$

$$\text{gdzie } \widehat{y_k} = \frac{B(q,\widehat{b})}{A(q,\widehat{a})} u_k \tag{14}$$

  uzyskane w danym kroku estymatory parametrów wykorzystuje się do generacji nowych zmiennych instrumentalnych, aż do ustabilizowania wyniku.

Jest znanym faktem w literaturze, że efektywność metody IV dla systemów liniowych jest silnie wrażliwa na zastosowane zmienne instrumentalne. Okazuje się, że w tym przypadku najbardziej właściwym w sensie minimalizacji asymptotycznej macierzy kowariancji estymatora (1) jest wybór macierzy Ψ_N różniący się od macierzy Φ_N tym, że zamiast rzeczywistych, zakłóconych (dostępnych dla pomiaru) wartości wyjść y_k zawierałaby ona wartości niezakłócone x_k (niestety pomiarowo niedostępne), tzn.

$$\psi_{k,opt.} = (-x_{k-1}, ..., -x_{k-n}, u_k, u_{k-1}, ..., u_{k-m})$$

gdzie x_k jest rozwiązaniem równania różnicowego dla $z_k = 0$

$$x_k = \frac{B(q, b^*)}{A(q, a^*)} u_k$$

Ponieważ wartości x_k są nieznane, a równocześnie brak jest w literaturze dowodu zbieżności metody iteracyjnej danej wzorem (13) – zrodził się pomysł, aby nieznane wielkości tego typu estymować metodami nieparametrycznymi, w szczególności opartymi na **jądrowej estymacji funkcji regresji**.

2. Optymalne zmienne instrumentalne

$$\begin{aligned} \Gamma_N &\triangleq \left(\frac{1}{N}\Psi_N^T \Phi_N\right)^{-1} \frac{1}{\sqrt{N}} \Psi_N^T \\ Z_N^* &\triangleq \frac{\frac{1}{\sqrt{N}} Z_N}{z_{\max}} \end{aligned}$$

$z_{\max}$ – górne ograniczenie

$$\Delta_N^{(IV)} = z_{\max} \Gamma_N Z_N^* \tag{15}$$

$$\|Z_N^*\|_2 = \sqrt{\sum_{k=1}^{N} \left(\frac{\frac{1}{\sqrt{N}} z_k}{z_{\max}}\right)^2} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left(\frac{z_k}{z_{\max}}\right)^2} \leq 1$$

$$Q(\Psi_N) = \max_{\|Z_N^*\|_2 \leq 1} \left\|\Delta_N^{(IV)}(\Psi_N)\right\|_2^2 \tag{16}$$

$\|\cdot\|_2$ – norma euklidesowa

Twierdzenie 1 *Wskaźnik jakości $Q(\Psi_N)$ dany wzorem (16) jest asymptotycznie optymalny dla*

$$\Psi_N^* = (\psi_1^*, \psi_2^*, ..., \psi_N^*)^T \qquad (17)$$
$$\psi_k^* = (x_{k-1}, ..., x_{k-n}, u_k, u_{k-1}, ..., x_{k-m})^T$$

gdzie $x_{k-1}, x_{k-2}, ..., x_{k-n}$ są wyjściami wolnymi od zakłóceń, tj.

$$\lim_{N\to\infty} Q(\Psi_N^*) \leqslant \lim_{N\to\infty} Q(\Psi_N) \quad \textit{z p. 1}$$

Dowód

Niech $z_{\max} = 1$

$$\left\| \Delta_N^{(IV)}(\Psi_N) \right\|^2 = \Delta_N^{(IV)T}(\Psi_N)\Delta_N^{(IV)}(\Psi_N) = Z_N^{*T}\Gamma_N^T\Gamma_N Z_N^*$$

$$Q(\Psi_N) = \max_{\|Z_N^*\|\leq 1}\left(\Delta_N^{(IV)T}(\Psi_N)\Delta_N^{(IV)}(\Psi_N)\right) = \max_{\|Z_N^*\|\leq 1}\left\langle Z_N^*, \Gamma_N^T\Gamma_N Z_N^*\right\rangle = \|\Gamma_N\|^2 = \lambda_{\max}\left(\Gamma_N^T\Gamma_N\right)$$

$\|\cdot\|$ – norma spektralna macierzy (indukowana przez normę euklidesową wektora)

$\lambda_{\max}()$ – największa wartość własna

$$\lambda_{\max}\left(\Gamma_N^T\Gamma_N\right) = \lambda_{\max}\left(\Gamma_N\Gamma_N^T\right)$$

wniosek

$$\begin{aligned}\max_{\|Z_N^*\|\leq 1}\left(\Delta_N^{(IV)T}(\Psi_N)\Delta_N^{(IV)}(\Psi_N)\right) &= \max_{\|\zeta\|\leq 1}\left\langle \zeta, \Gamma_N\Gamma_N^T\zeta\right\rangle \\ &= \max_{\|\zeta\|\leq 1}\left\langle \zeta, \left(\frac{1}{N}\Psi_N^T\Phi_N\right)^{-1}\left(\frac{1}{N}\Psi_N^T\Psi_N\right)\left(\frac{1}{N}\Phi_N^T\Psi_N\right)^{-1}\zeta\right\rangle\end{aligned}$$

dla $N \to \infty$ zachodzi z p. 1

$$\max_{\|Z_N^*\|\leq 1}\left(\Delta_N^{(IV)T}(\Psi_N)\Delta_N^{(IV)}(\Psi_N)\right) = \max_{\|\zeta\|\leq 1}\left\langle \zeta, \left(\frac{1}{N}\Psi_N^T\Psi_N^{*T}\right)^{-1}\left(\frac{1}{N}\Psi_N^T\Psi_N\right)\left(\frac{1}{N}\Psi_N^{*T}\Psi_N\right)^{-1}\zeta\right\rangle$$

Lemat 1 *(techniczny) Niech M_1 i M_2 będą macierzami o tych samych wymiarach. Jeśli istnieją $\left(M_1^T M_1\right)^{-1}$, $\left(M_1^T M_2\right)^{-1}$ i $\left(M_2^T M_1\right)^{-1}$, to macierz*

$$D_N = \left(M_2^T M_1\right)^{-1} M_2^T M_2 \left(M_1^T M_2\right)^{-1} - \left(M_1^T M_1\right)^{-1}$$

jest nieujemnie określona, tzn. dla każdego wektora ζ zachodzi

$$\zeta^T D_N \zeta \geq 0$$

Podstawiając $M_1 = \frac{1}{\sqrt{N}} \Psi_N^{*T}$, $M_2 = \frac{1}{\sqrt{N}} \Psi_N$

$$\zeta^T \Gamma_N \Gamma_N^T \zeta \geq \zeta^T \left(\frac{1}{N} \Psi_N^{*T} \Psi_N^*\right)^{-1} \zeta$$

z p.1 dla każdego wektora ζ. Zatem

$$Q\left(\Psi_N\right) = \max_{\|\zeta\| \leq 1} \left(\zeta^T \Gamma_N \Gamma_N^T \zeta\right) \geq \max_{\|\zeta\| \leq 1} \left(\zeta^T \left(\frac{1}{N} \Psi_N^{*T} \Psi_N^*\right)^{-1} \zeta\right) \tag{18}$$

Ponieważ dla $\Psi_N = \Psi_N^*$ obie strony w (18) są równe, $Q\left(\Psi_N\right)$ osiąga infimum dla $\Psi_N = \Psi_N^*$.