ALEKSANDER ZELIAŚ

PREDYKCJA SZEREGÓW CZASOWYCH W WARUNKACH AUTOKORELACJI SKŁADNIKÓW LlosowYCH

Przyjmijmy, że rozpatrujemy liniowy model ekonometryczny postaci

$$Y_t=\beta_1 X_{t1}+\beta_2 X_{t2}+\ldots+\beta_k X_{tk}+u_t \quad (t=1, \ldots, n), \tag{1}$$

gdzie:

Y_t — zmienna endogeniczna, wyjaśniana przez dane równanie,

X_{tj} — zmienne objaśniające, występujące w postaci n zbiorów w powtarzanych próbach, z czego zmienna X_{t1} przybiera stale wartość 1,

β_j — parametry strukturalne,

u_t — składnik losowy.

Każde z n układów równań (1) można zapisać w postaci

$$\mathbf{y}=\mathbf{X\beta}+\mathbf{u}, \tag{2}$$

gdzie symbol **y** oznacza wektor *(nxl)* zaobserwowanych wartości zmiennej endogenicznej Y_t, X jest macierzą *(nxk)* zaobserwowanych wartości zmiennych objaśniających X_{tj} *(j*—1, ..., *k),* zaś ß oznacza wektor *(kx1)* nieznanych a priori wartości parametrów strukturalnych β_i, a u niech oznacza wektor (*nx1*) składników losowych u_t.

Zakładając, że

1° $E(\mathbf{u})=0$,

2° $E(\mathbf{uu}')=\sigma_u^2 \mathbf{I}_n$, gdzie I_n jest macierzą jednostkową o wymiarach *nxn,*

3° **X** jest macierzą ustalonych wartości zmiennych objaśniających X_{tj} *(j=1, ...,k)*,

4° **X** jest rzędu $r(\mathbf{X})=k<n$,

5° Składnik losowy u_t jest nie skorelowany ze zmiennymi objaśniającymi,

nieznane paramery βj *(j=1, ...,k)* szacuje się na podstawie klasycznej metody najmniejszych kwadratów.

W pewnych sytuacjach może się zdarzyć (najczęściej gdy obserwacje mają postać szeregów czasowych), że składniki losowe $u_1, \ldots, u_n$ rozpatrywanego modelu nie są od siebie stochastycznie niezależne. Zależność

taką nazywa się zwykle w literaturze autokorelacją składnika losowego (ang. autocorrelation of the disturbance term).

W niniejszej pracy zajmować się będziemy zasadniczo sytuacją, gdy przestaje być prawdziwe założenie:

$$cov(u_t, u_s)=0 \quad \text{dla} \quad s\neq t,$$

natomiast pozostałe pozostają w mocy. W przypadkach, gdy istnieje autokorelacja składników losowych, przyjmuje się zwykle, że zależność między nimi może być przedstawiona za pomocą relacji liniowej o postaci ogólnej

$$u_t=\sum_{i=1}^{p}\gamma_i u_{t-i}+\varepsilon_t, \tag{3}$$

gdzie p jest liczbą naturalną, γ_i są stałymi parametrami, a ε_t jest procesem czysto losowym o nadziei matematycznej równej zero i skończonej wariancji.

Parametry γ_i *(i=1, ..., p)* są nieznane a priori, ale możliwe jest znalezienie metod, za pomocą których można otrzymać estymatory zgodne. Metodą taką jest w szczególności metoda najmniejszych kwadratów.

W przypadku gdy *p=1* (co ma często miejsce w praktyce) relacja (3) ulegnie uproszczeniu i przyjmie postać

$$u_t=\gamma_1 u_{t-1}+\varepsilon_t, \tag{4}$$

przy czym $|\gamma| < 1$, a ε_t jest procesem czysto losowym.

Ciąg zmiennych losowych $\{\varepsilon_t\}$ spełnia oczywiście założenia:

$$E(\varepsilon_t)=0,$$

$$E[\varepsilon_t-E(\varepsilon_t)]^2=E(\varepsilon_t)^2=\sigma_\varepsilon^2,$$

$$cov(\varepsilon_t, \varepsilon_s)=E\{[\varepsilon_t-E(\varepsilon_t)][\varepsilon_s-E(\varepsilon_s)]\}=0 \quad \text{dla} \quad s\neq t.$$

Można wykazać[1], że skorelowany w czasie składnik losowy u_t *(t=1, ..., n)* ma nadzieję matematyczną zero i stałą (niezależną od wskaźnika bieżącego *t*) wariancją σ_u^2, o wartości skończonej. Oznacza to, że

$$E(u_t)=0,$$

$$\sigma_{u_t}^2=\frac{\sigma_\varepsilon^2}{1-\gamma_1^2}=\sigma_u^2 \quad \text{dla każdego } t, \tag{5}$$

a tym samym u_t jest procesem stacjonarnym[2].

[1] Por. monografię J. Johnstona, *Econometric methods,* 2 nd Edition, New York 1972, s. 243 - 246.

[2] Procesem stacjonarnym jest taki proces stochastyczny, którego charakterystyki nie zależą od parametru *t*. Dokładniej mówiąc, wszystkie zmienne losowe składające się na proces stacjonarny mają ten sam rozkład o tych samych wartościach parametrów. Por. np. pracę Z. Pawłowskiego, *Wstęp do statystyki matematycznej,* Warszawa 1965, s. 378 - 385.

Łatwo sprawdzić, że kowariancja składników losowych $u_1, \ldots, u_n$ jest równa

$$E(u_t u_{t-s}) = (\gamma_1)^s \sigma_u^2 \tag{6}$$

Widzimy w ten sposób, że równanie (1) nie spełnia założenia 2°. Wnioskujemy zatem, że składnik losowy u_t jest skorelowany w czasie, co odpowiada sytuacji, że występujący w (4) parametr $\gamma_1 \neq 0$.

Z równania (6) wynika dalej, że jest

$$\frac{E(u_t u_{t-s})}{\sigma_u^2} = (\gamma_1)^s \quad \text{dla} \quad s = 0, 1, \ldots \tag{7}$$

Wyrażenie występujące po lewej stronie wzoru (7) jest współczynnikiem autokorelacji rządu s, który oznaczamy przez ϱ_S. Mierzy on liniową zależność między zmiennymi losowymi u_t o wskaźnikach t różniących się od siebie o s jednostek. Stąd (7) możemy teraz wyrazić za pomocą wzoru

$$\rho_s = (\gamma_1)^s \quad \text{dla} \quad s = 0, 1, \ldots,$$

co oznacza, że kolejne współczynniki autokorelacji wyrażają się jako kolejne potęgi parametru γ_1.

Reasumując wyniki tej dyskusji można stwierdzić, że w przypadku, gdy mamy relację postaci (4) i gdy $|\gamma_1| < 1$, macierz $E(\mathbf{uu}')$ nie jest już, macierzą diagonalną, ale posiada różne od zera elementy także i poza główną przekątną. Macierz $E(\mathbf{uu}')$ (o wymiarach nxn) jest wówczas postaci

$$E(\mathbf{uu}') = \mathbf{V} = \sigma_u^2 \begin{bmatrix} 1 & \gamma_1 & \gamma_1^2 & \ldots & \gamma_1^{n-1} \\ \gamma_1 & 1 & \gamma_1 & \ldots & \gamma_1^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_1^{n-1} & \gamma_1^{n-2} & \gamma_1^{n-3} & \ldots & 1 \end{bmatrix} = \sigma_u^2 \boldsymbol{\Omega}, \tag{8}$$

przy czym zakładamy, że elementy macierzy $\boldsymbol{\Omega}$ są znane oraz $tr\boldsymbol{\Omega} = n$.

Dowodzi się [3], że jeżeli zmienne objaśniające są nielosowe, wówczas istnienie autokorelacji składników losowych nie wpływa na nieobciążoność lub zgodność estymatorów parametrów strukturalnych i własności te pozostają takie, jak w przypadku ich niezależności. Okazuje się jednak, że autokorelacja zmienia wariancje i kowariancje tych estymatorów (wariancje estymatorów parametrów strukturalnych otrzymywane za pomocą metody najmniejszych kwadratów będą większe niż w przypadku, gdy nie istnieje autokorelacja składnika losowego).

W literaturze ekonometrycznej spotkać można wiele prac poświęconych metodom estymacji parametrów modelu (2) w przypadku istnienia autokorelacji składników losowych[4]. Jedną z takich metod, mogącą mieć

[3] Por. J. Johnston, op. cit., s. 246.

[4] A. C. Aitken, *On least squares and linear combination of observations,* Proceedings of the Royal Society of Edinburgh, 1935, vol. 55; D. Cochrane, G. H. Orcutt, *Application of least squares regression to relationships containing autocorrelated*

zastosowanie w przypadku, gdy między kolejnymi składnikami losowymi u_t zachodzą relacje postaci (4) jest tak zwana uogólniona metoda najmniejszych kwadratów Aitkena [5], uwzględniająca niediagonalny charakter macierzy E (**uu'**).

Przy stosowaniu uogólnionej metody najmniejszych kwadratów parametry modelu (2) wyznacza się z równania macierzowego

$$\mathbf{b}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}=(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}, \tag{9}$$

przy czym b jest wektorem ocen odpowiadającym uogólnionej metodzie najmniejszych kwadratów, podczas gdy macierz $\boldsymbol{\Omega}^{-1}$ ma postać

$$\boldsymbol{\Omega}^{-1}=\frac{1}{1-\gamma_1^2}\begin{bmatrix} 1 & -\gamma_1 & 0 & \dots & 0 & 0 & 0 \\ -\gamma_1 & 1+\gamma_1^2 & -\gamma_1 & \dots & 0 & 0 & 0 \\ 0 & -\gamma_1 & 1+\gamma_1^2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & & -\gamma_1 & 1+\gamma_1^2 & -\gamma_1 \\ 0 & 0 & 0 & & 0 & -\gamma_1 & 1 \end{bmatrix}$$

Macierz wariancji i kowariancji tych estymatorów dana jest wzorem

$$\begin{aligned} var(\mathbf{b}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}= \\ &= \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \end{aligned} \tag{10}$$

gdzie σ_u^2 (na mocy (5) $\sigma_u^2=\sigma_\varepsilon^2/(1-\gamma_1^2)$ jest wariancją zmiennej losowej u_t.

Przeprowadzając odpowiednie obliczenia otrzymujemy

$$\mathbf{y}=\mathbf{X}\mathbf{b}+\mathbf{e}, \tag{11}$$

gdzie wektor reszt **e=y—Xb .**

Przyjmijmy teraz, że *yr* oznacza prognozę dla zmiennej endogenicznej *Y* w okresie prognozowanym *T=n+1, n+2, ..., n+s,* otrzymaną na podstawie wyrażenia

$$y_T=\mathbf{x}_T\boldsymbol{\beta}+u_T, \tag{12}$$

gdzie $\mathbf{x}_T$ oznacza wektor wierszowy ustalonych wartości zmiennych objaśniających modelu w okresie prognozowanym *T*, tak że [6]

$$\mathbf{x}_T=[x_{T1}\,x_{T2}\dots x_{Tk}],$$

error terms, Journal of the American Statistical Association, 1949, vol. 44; K. R. Kadiyala, *A trasformation used to circumvent the problem of autocorrelation,* Econometrica, 1968, vol. 36.

[5] Por. A. C. Aitken, op. cit.

[6] Przewidywane wartości zmiennych objaśniających X_{Tj} *(j=1, ..., k)* dla okresu prognozowanego można ustalić na poziomie planowanym, a jeśli brak jest planu, to najsłuszniej jest je ustalić poprzez wyznaczenie, a następnie ekstrapolację trendów wartości tych zmiennych.

a u_T jest składnikiem losowym w okresie prognozowanym T.

Przyjmijmy również, że

$$E(u_T)=0,$$

$$E(u_T)^2=\sigma_T^2=\sigma^2,$$

$$E(\mathbf{u}u_T)=\begin{bmatrix} E(u_1 u_T) \\ E(u_2 u_T) \\ \vdots \\ E(u_n u_T) \end{bmatrix}=\mathbf{w},$$

gdzie **w** jest wektorem *(nx1)* kowariancji składników losowych $u_1, \ldots, u_n$ ze składnikiem losowym u_T .

Wyrażenie

$$p=\mathbf{c}'\mathbf{y} \tag{13}$$

nazywać będziemy liniowym predyktorem, w którym **c** oznacza wektor *(nx1)* odpowiednio ustalonych wag.

Jeżeli nieobciążony predyktor p ma być najbardziej efektywny[7], to wagi **c** dobieramy tak, aby zminimalizować wariancję predykcji

$$\sigma_p^2=E\{(p-y_T)^2\},$$

przy czym $E(p-y_T)=0$.

Wykorzystując (13) i (12) mamy

$$\begin{aligned} p-y_T&=\mathbf{c}'\mathbf{y}-\mathbf{x}_T\boldsymbol{\beta}-u_T= \\ &=\mathbf{c}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{u})-\mathbf{x}_T\boldsymbol{\beta}-u_T= \\ &=(\mathbf{c}'\mathbf{X}-\mathbf{x}_T)\boldsymbol{\beta}+\mathbf{c}'\mathbf{u}-u_T \end{aligned}$$

Nieobciążoność predyktora p wymaga, by wektor **c** spełniał następującą równość

$$\mathbf{c}'\mathbf{X}-\mathbf{x}_T=\mathbf{0}, \tag{14}$$

a zatem błąd predykcji jest równy

$$p-y_T=\mathbf{c}'\mathbf{u}-u_T.$$

Stąd wariancja predykcji — jak to już wcześniej zasygnalizowano — jest równa

$$\begin{aligned} \sigma_p^2&=E\{(p-y_T)^2\}= \\ &=E\{(p-y_T)(p-y_T)'\}= \end{aligned}$$

[7] Predyktor nazywa się najbardziej efektywny, jeżeli jest nieobciążony i przy każdym n wariancja jego jest mniejsza od wariancji innych możliwych predyktorów przy tej samej liczbie obserwacji n. Por. w tej sprawie pracę Z. Pawłowskiego, *Prognozy ekonometryczne,* Warszawa 1973.

$$= E\{(\mathbf{c}'\mathbf{u} - u_T)(\mathbf{c}'\mathbf{u} - u_T)'\} =$$
$$= E\{(\mathbf{c}'\mathbf{u} - u_T)(\mathbf{u}'\mathbf{c} - u_T')\} =$$
$$= E\{\mathbf{c}'\mathbf{u}\mathbf{u}'\mathbf{c} + u_T^2 - 2\mathbf{c}'\mathbf{u}u_T\},$$

skąd dostajemy ostatecznie

$$\sigma_p^2 = \mathbf{c}'\mathbf{V}\mathbf{c} + \sigma_T^2 - 2\mathbf{c}'\mathbf{w} \tag{15}$$

Minimalizujemy teraz funkcję określoną wzorem (15) przy uwzględnieniu (14)r Mamy w ten sposób:

$$\varphi = \mathbf{c}'\mathbf{V}\mathbf{c} - 2\mathbf{c}'\mathbf{w} - 2(\mathbf{c}'\mathbf{X} - \mathbf{x}_T)\boldsymbol{\lambda}, \tag{16}$$

gdzie $\boldsymbol{\lambda}$ jest wektorem *(kx1)* mnożników Lagrange'a.

Łatwo można wykazać, że wykorzystanie warunku koniecznego na istnienie ekstremum funkcji $\varphi\left(\frac{\partial\varphi}{\partial\mathbf{c}} = \mathbf{0},\ \frac{\partial\varphi}{\partial\boldsymbol{\lambda}} = \mathbf{0}\right)$ prowadzi do układu równań:

$$2\mathbf{V}\mathbf{c} - 2\mathbf{w} - 2\mathbf{X}\boldsymbol{\lambda} = \mathbf{0},$$
$$-2\mathbf{c}'\mathbf{X} + 2\mathbf{x}_T = \mathbf{0}.$$

Ostatecznie więc mamy:

$$\begin{aligned} \mathbf{V}\mathbf{c} - \mathbf{X}\boldsymbol{\lambda} &= \mathbf{w}, \\ \mathbf{X}'\mathbf{c} &= \mathbf{x}_T' \end{aligned} \tag{17}$$

Zauważmy, że układ równań (17) można zapisać w postaci macierzowej. Mamy wtedy

$$\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ -\hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{w} \\ \mathbf{x}_T' \end{bmatrix} \tag{18}$$

Poszukiwany wektor $\begin{bmatrix} \hat{\mathbf{c}} \\ -\hat{\boldsymbol{\lambda}} \end{bmatrix}$ znajdujemy rozwiązując układ (18), względem $\mathbf{c}$ i $-\hat{\boldsymbol{\lambda}}$. Ostatecznie więc mamy

$$\begin{bmatrix} \hat{\mathbf{c}} \\ -\hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{w} \\ \mathbf{x}_T' \end{bmatrix} \tag{19}$$

Podmacierze **V**, **X**, **X'** i podmacierz zerowa **0** mają wymiary *nxn, nxk, kxn* oraz *kxk,* a wektory **c**, —.$\hat{\boldsymbol{\lambda}}$, **w** i $\mathbf{x}'_{\mathbf{T}}$ mają wymiary *nx1, kx1, nx1* oraz *1xk.*

Wykorzystując regułę [8] o wyznaczaniu macierzy odwrotnej $\mathbf{A}^{-1}$ przez rozkład macierzy kwadratowej **A** (która w naszym przypadku ma wymiary (n+k x(n+k)) na podmacierze (bloki) **V**, **X**, **X** i **0**, znajdujemy ma-

[8] H. E. Kryński, *Matematyka dla ekonomistów,* Warszawa 1973, wyd. 5, s. 292 - - 297.

cierz odwrotną $\mathbf{A}^{-1}$, a następnie wyznaczamy poszukiwany wektor wag $\hat{\mathbf{c}}$.

Zatem mamy

$$\begin{bmatrix} \hat{\mathbf{c}} \\ -\hat{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{V}^{-1}+\mathbf{V}^{-1}\mathbf{X}(\mathbf{0}-\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} & -\mathbf{V}^{-1}\mathbf{X}(\mathbf{0}-\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ -(\mathbf{0}-\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} & (\mathbf{0}-\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{w} \\ \mathbf{x}'_T \end{bmatrix}, \tag{20}$$

skąd

$$\hat{\mathbf{c}} = \mathbf{V}^{-1}[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{w}+\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T$$

W rezultacie, najlepszy nieobciążony predyktor ma w naszym przykładzie postać

$$\begin{aligned}
\hat{p} &= \hat{\mathbf{c}}'\mathbf{y} = \\
&= \{\mathbf{V}^{-1}[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{w}+\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T\}\mathbf{y} = \\
&= \{\mathbf{w}'[\mathbf{I}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}']\mathbf{V}^{-1}+\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\}\mathbf{y} = \\
&= \mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}+\mathbf{w}'\mathbf{V}^{-1}\mathbf{y}-\mathbf{w}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'^{-1}\mathbf{y} = \\
&= \mathbf{x}_T\mathbf{b}+\mathbf{w}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b}),
\end{aligned}$$

skąd dostajemy ostatecznie [9]

$$\hat{p} = \mathbf{x}_T\mathbf{b}+\mathbf{w}'\mathbf{V}^{-1}\mathbf{e} \tag{21}$$

A. S. Goldberger[10] wykazał, że wariancja predykcji odpowiadająca predyktorowi (21) jest równa

$$\sigma^2_{\hat{p}} = \sigma^2_T - 2\hat{\mathbf{c}}'\mathbf{w}+\hat{\mathbf{c}}'\mathbf{V}\hat{\mathbf{c}}$$

Wykorzystując, że

$$\hat{\mathbf{c}} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T+\mathbf{V}^{-1}[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{w},$$

możemy napisać

$$\begin{aligned}
\sigma^2_{\hat{p}} = \sigma^2_T &- 2\{\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}+\mathbf{w}'[\mathbf{I}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}']\mathbf{V}^{-1}\}\mathbf{w}+ \\
&+\{\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}+\mathbf{w}'[\mathbf{I}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}']\mathbf{V}^{-1}\}\mathbf{V}\times \\
&\times\{\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T+\mathbf{V}^{-1}[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{w}\},
\end{aligned}$$

skąd dostajemy

$$\begin{aligned}
\sigma^2_{\hat{p}} = \sigma^2_T &+ \mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T-2\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{w}- \\
&-\mathbf{w}'[\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{w}.
\end{aligned} \tag{22}$$

[9] Por. pracę A. S. Goldberga, *Best linear unbiased prediction in the generalized linear regression model,* Journal of the American Statistical Association, 1962, vol. 57, ss. 369 - 375. Dokładniej mówiąc, A. S. Goldberger podał sposób modyfikacji predyktorów klasycznych pozwalający na uwzględnienie w procesie predykcji informacji o charakterze związku autokorelacyjnego.

[10] A. S. Goldberger, op. cit.

Jak się okazuje, wyrażenie (22) jest zawsze mniejsze od wariancji predykcji w przypadku, gdy posługujemy się predyktorem typu klasycznego, a więc wyrażeniem

$$\overline{p}=\mathbf{x}_T\mathbf{b}$$

Wariancja predykcji odpowiadająca predyktorowi $\overline{p}$ — podobnie jak przedtem jest równa

$$\sigma^2_{\overline{p}}=\sigma^2_T-2\overline{\mathbf{c}}'\mathbf{w}+\overline{\mathbf{c}}'\mathbf{V}\overline{\mathbf{c}}$$

Z (20) otrzymujemy, że

$$\overline{\mathbf{c}}=\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T$$

Tak więc

$$\sigma^2_{\overline{p}}=\sigma^2_T-2\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{w}+\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\times$$
$$\times\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T=$$
$$=\sigma^2_T-2\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{w}+\mathbf{x}_T(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}'_T \tag{23}$$

Porównanie (23) z (22) pokazuje, że

$$\sigma^2_{\overline{p}}=\sigma^2_{\hat{p}}+\mathbf{w}'[\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{w} \tag{24}$$

Z powyższego wynika, że wariancja klasycznego predyktora $\sigma^2_{\overline{p}}$ jest większa od wariancji predykcji $\sigma^2_{\hat{p}}$. Istotnie wykorzystując (24) otrzymujemy, że rząd wielkości zysku na efektywności jest równy:

$$\mathbf{w}'[\mathbf{v}^{-1}-\mathbf{v}^{-1}\mathbf{x}(\mathbf{x}'\mathbf{v}^{-1}\mathbf{x})^{-1}\mathbf{x}'\mathbf{v}-1]\mathbf{w}.$$

Korzystając ze wzoru (21) stwierdzamy, że prognoza $\hat{y}_T$ przy użyciu predyktora $\hat{p}$ jest równa.

$$\hat{y}_T=\mathbf{x}_T\mathbf{b}+\mathbf{w}'\mathbf{V}^{-1}\mathbf{e}\quad(T=n+1), \tag{25}$$

gdzie **b** jest wektorem ocen parametrów strukturalnych otrzymanych za pomocą uogólnionej metody najmniejszych kwadratów Aitkena, **w** jest wektorem *(nx1)* kowariancji składników losowych $u_1,\ldots,u_n$ ze składnikiem losowym w_{n+1}, V jest macierzą *(nxn)* wariancji i kowariancji składników losowych modelu, **e** jest wektorem *(nx1)* reszt modelu, czyli

$$\mathbf{w}=\begin{bmatrix}E(u_1u_{n+1})\\E(u_2u_{n+1})\\\vdots\\E(u_nu_{n+1})\end{bmatrix},\quad \mathbf{e}=\begin{bmatrix}e_1\\e_2\\\vdots\\e_n\end{bmatrix},\quad \mathbf{V}=\sigma^2_u\begin{bmatrix}1&\gamma_1&\gamma_1^2&\ldots&\gamma_1^{n-1}\\\gamma_1&1&\gamma_1&\ldots&\gamma_1^{n-2}\\\ldots&\ldots&\ldots&\ldots&\ldots\\\gamma_1^{n-1}&\gamma_1^{n-2}&\gamma_1^{n-3}&\ldots&1\end{bmatrix}.$$

Ponieważ istniejące powiązanie składników losowych spełnia relację: $u_t=\gamma_i u_{t-1}+\varepsilon_t$, zatem ze wzoru (6) wynika oczywista równość

$$\mathbf{w}=\begin{bmatrix}\sigma_u^2&\gamma_1^n\\\sigma_u^2&\gamma_1^{n-1}\\&\vdots\\\sigma_u^2&\gamma_1\end{bmatrix}=\gamma_1\sigma_u^2\begin{bmatrix}\gamma_1^{n-1}\\\gamma_1^{n-2}\\\vdots\\1\end{bmatrix}.$$

Jak łatwo zauważyć, współrzędne wektora $\mathbf{w}$, po pomnożeniu przez γ_1, odpowiadają elementom występującym w ostatniej kolumnie macierzy $\mathbf{V}$.

Ponieważ $\mathbf{V}\mathbf{V}^{-1}=\mathbf{I}_n$, przeto $\mathbf{V}'\mathbf{V}^{-1}$, po przemnożeniu przez γ_1, jest równe ostatniemu wierszowi macierzy jednostkowej $\mathbf{I}_n$, tak że

$$\mathbf{w}'\mathbf{V}^{-1}=\gamma_1[0 \;\; 0 \;\; \ldots \;\; 1].$$

W rezultacie

$$\mathbf{w}'\mathbf{V}^{-1}\mathbf{e}=\gamma_1 e_n.$$

Poszukiwana prognoza ma zatem postać ostateczną

$$\hat{y}_{n+1}=\mathbf{x}_{n+1}\mathbf{b}+\gamma_1 e_n \tag{26}$$

Jak łatwo zauważyć, pierwszy składnik prawej strony wzoru (26) jest prognozą klasyczną. Natomiast drugi składnik $\gamma_1 e_n$ jest poprawką ze względu na autokorelację.

Budowa prognozy na okres $T=n+s$ przebiega analogicznie jak poprzednio. Mamy zatem

$$\hat{p}=\mathbf{x}_T\mathbf{b}+\mathbf{w}'\mathbf{V}^{-1}\mathbf{e}, \tag{27}$$

gdzie:

$$\mathbf{w}=E(uu_{n+s})=\begin{bmatrix}E(u_1u_{n+s})\\E(u_2u_{n+s})\\\vdots\\E(u_nu_{n+s})\end{bmatrix}=\begin{bmatrix}\sigma_u^2 & \gamma_1^{n-1+s}\\\sigma_u^2 & \gamma_1^{n-2+s}\\ & \vdots\\\sigma_u^2 & \gamma_1^{s}\end{bmatrix}=(\gamma_1)^s\sigma_u^2\begin{bmatrix}\gamma_1^{n-1}\\\gamma_1^{n-2}\\\vdots\\1\end{bmatrix}$$

W konsekwencji (27) możemy teraz przedstawić w postaci

$$\hat{p}=\mathbf{x}_T\mathbf{b}+(\gamma_1)^s[0 \;\; 0 \;\; \ldots \;\; 1]\mathbf{e}=\mathbf{x}_T\mathbf{b}+(\gamma_1)^s e_n$$

Otrzymaliśmy więc, że prognoza $\hat{y}_T$ *(T=n+s)* przy użyciu predyktora $\hat{p}$ jest równa

$$\hat{y}_{n+s}=\mathbf{x}_{n+s}\mathbf{b}+(\gamma_1)^s e_n \tag{28}$$

Dla $s=1$, wyrażenie (28) przybiera postać (26).

Stosowanie tej metody predykcji do praktycznych obliczeń wymaga znajomości parametru γ_1. Wobec tego zwykle szacuje się go na podstawie próby [11]. Należy jednak pamiętać, że z uwagi na to, iż γ_1 jest oceną parametru γ_1, a więc różniąca się nieco od γ_1, relacja (28) będzie spełniona tylko w przybliżeniu.

[11] Gdy próba statystyczna służąca do estymacji modelu jest mała *(n<30)* stwarza to niebezpieczeństwo, że parametr γ_1 będzie źle oszacowany, a w konsekwencji będzie zwykle prowadził do znacznych błędów predykcji.

PREDICTION PROBLEMS IN THE SITUATION WHEN DISTURBANCES ARE AUTOCORRELATED IN THE LINEAR MODEL

Summary

In the article the autor discusses prediction problems in the general linear model where disturbances are autocorrelated. Suppose we have sample data for periods 1 to n from the model $\mathbf{y}=\mathbf{X}\beta+\mathbf{u}$ with $E(\mathbf{u})=\mathbf{0}$ and $E(\mathbf{uu}')=\mathbf{V}$. Suppose further that we are given the row vector $\mathbf{x}_{n+1}$ of values of the explanatory variables in period $n+1$. The problem now how to estimate y_{n+s} when $E(\mathbf{uu}')=\mathbf{V}=\sigma^2_u\mathbf{\Omega}$, where $\mathbf{\Omega}$ is assumed to be known, symmetric, positivedefinite matrix. If the disturbances follows a first-order scheme $u_t=\gamma_1 u_{t-1}+\varepsilon_t$ where $|\gamma_1|<1$, then we can see that the best linear unbiased predictor is $\hat{y}_{n+1}=\mathbf{x}_{n+s}\mathbf{b}+(\gamma_1)^s\cdot e_n$ where $\mathbf{b}$ is the generalized least-squares estimator and e_n is the last row of the vector $\mathbf{e}=\mathbf{y}-\mathbf{Xb}$ of generalized least-squares residuals.

When the elements of $\mathbf{\Omega}$ (that is, the value of γ_1) were unknown formula pictured above should still be used for prediction purposes with $\mathbf{b}$ and γ_1 replaced by estimates.

ESTIMATION PROBLEMS OF PARAMETERS OF ECONOMETRIC MODELS WITH DISTRIBUTED LAGS

Summary

Let us assume that a linear relationship exists between a variable Y_t and $k-1$ explanatory variables X_t, X_{t-1}, ..., X_{t-k} and a disturbance term u. If we have a sample of n observations on Y and $X's$ we can write

$$Y_t=L+\sum_{j=0}^{k}\beta_j x_{t-j}+u_t \qquad (t=1, \ldots, n) \tag{1}$$

The α, $\beta_0-\beta_1$, ..., β_k coefficients and the parameters of the u distribution are unknown and our problem is to obtain estimates of these unknowns. Under the usual assumption about the distribution of u and the independence of X and u there are in principle no new estimation problems in this model. Least-squares will give best linear unbiased estimates, if the model has been specified, correctly. Several difficulties, however, are likely to arise in practice. First of all one cannot really expect any precise and firm indication from theory of the length of lag to be incorporated; rather one hopes to determine the lag from the data by fitting a fairly long lag and then examining the significance of the coefficients of various lagged values of X. But this in turn raises two main statistical difficulties; one is that observations are lost due to the lags and the other is that typically the various lagged values of X will be highly intercorrelated leading to very imprecise estimates of the lagged coefficients and great difficulty in making useful inferences about them.

These difficulties have lead to the a priori imposition of various assumptions about the form of the weights β_0, ..., β_k in an attempt to produce a more amenable estimation problem involving fewer than $k+1$ parameters. The above paper discusses three models for distributed-lag analysis that either reduce the number of observations lost due to lagging and/or reduce the number of parameters to be estimated.