

Stanovení chyby předpovědi

Uvažujme opět standardní normální lineární regresní model, u kterého jsme již předtím ověřili jeho způsobilost k prognózování minimálně ve smyslu ověření stability modelové struktury.

Budeme pracovat s modelem v zápisu (pro predikované období o délce m)

$$(1) \quad y_p = X_p \cdot \beta + \varepsilon_p$$

y_p je [m -členný] sloupcový vektor hodnot vysvětlované (endogenní) proměnné, kterých nabude v predikovaném období

X_p je [$m \times k$ - rozměrná] matice hodnot vysvětlujících (exogenních) proměnných realizovaných v predikovaném období

β je [k -členný] sloupcový vektor regresních parametrů (hodnoty parametrů přejímáme beze změn z pozorovaného období)

ε_p je [m -členný] sloupcový vektor náhodných složek predikovaného období

Předpokládáme přitom, že

$$a) \quad E(\varepsilon) = 0_T, \quad E(\varepsilon_p) = 0_m,$$

tzn. že náhodné složky s nulovou střední hodnotou v pozorovaném období zůstanou centrované i v předpovídaném období.

$$b) \quad \text{Var}(\varepsilon_p^{(j)}) = \text{Var}(\varepsilon^{(i)}) = \sigma^2, \quad i = 1, 2, \dots, T; \quad j = 1, 2, \dots, m;$$

tzn. že rozptyly náhodných složek pozorovaného i predikovaného období jsou shodné (rovné σ^2).

$$c) \quad E(\varepsilon_p^{(j)} \cdot \varepsilon^{(i)}) = 0, \quad i = 1, 2, \dots, T; \quad j = 1, 2, \dots, m;$$

tzn. že náhodné složky predikovaného období jsou nekorelované navzájem i s náhodnými složkami pozorovaného období. Souhrnná kovarianční matice $\tilde{\Sigma}_{[T+m; T+m]}$ má tedy tvar

$$\tilde{\Sigma} = \text{Cov} \begin{pmatrix} \varepsilon \\ \varepsilon_p \end{pmatrix} = \begin{pmatrix} \sigma^2 I_T & 0 \\ 0 & \sigma^2 I_m \end{pmatrix}$$

d) Náhodné složky jsou nekorelované s vysvětlujícími proměnnými v pozorovaném i v predikovaném období, tj. $E(X_j' \varepsilon) = 0$ pro $j = 1, 2, \dots, T$

$$E(X_p' \varepsilon_p) = 0 \quad \text{pro } l = 1, 2, \dots, m.$$

Model můžeme přehledněji znázornit maticovým zápisem

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_T \\ y_{T+1} \\ \dots \\ y_{T+m} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ x_{T,1} & x_{T,2} & x_{T,3} & \dots & x_{T,k} \\ x_{T+1,1} & x_{T+1,2} & x_{T+1,3} & \dots & x_{T+1,k} \\ \dots & \dots & \dots & \dots & \dots \\ x_{T+m,1} & x_{T+m,2} & x_{T+m,3} & \dots & x_{T+m,k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_T \\ \varepsilon_{T+1} \\ \dots \\ \varepsilon_{T+m} \end{pmatrix}.$$

Bodovou předpověď m -členného vektoru závisle proměnné získáme pomocí **predikční funkce**¹ tvaru

$$(2) \quad y_p^* = X_p \cdot b = X_p \cdot (X'X)^{-1} X'y \quad , \text{ kde}$$

X je $[Txk - \text{rozměrná}]$ matice pozorovaných hodnot k vysvětlujícím proměnných

Predikční funkce 2), jež je zřejmě lineární funkcí y , je nestrannou funkcí vektoru předpovědí y_p v tom smyslu, že pro ni platí

$$(3) \quad E(y_p^*) = E(X_p \cdot b) = X_p E(b) = X_p \cdot \beta = E(y_p) \quad .$$

[X_p je nestochastická] [b je nestranný odhad β]

To znamená, že střední hodnota predikční funkce je rovna střední hodnotě y_p ²

Poznámka Neznačí to však, že $E(y_p^*) = y_p$ ³

Lze ukázat, že predikční funkce 2) má ze všech nestranných predikčních funkcí (lineárními v y) nejmenší v maticovém smyslu kovarianční matici vektoru reziduí v období předpovědi, tj. následujícího vektoru d_p

Vektor chyb d_p definujeme jako odchylku (chybu) předpovídané hodnoty y_p^* od skutečné y_p :

$$(4) \quad d_p = y_p^* - y_p = X_p(b - \beta) - \varepsilon_p$$

Všimněme si, že zdrojem chyb předpovědi vektoru skutečných hodnot y_p je jednak variabilita náhodné složky ε_p , jednak výběrová chyba odhadové funkce b zpravidla pořízené *metodou nejmenších čtverců OLS*:

Střední chyba vektoru chyb predikčního vektoru d_p v (4) je nulová, protože platí

¹ Predikční funkce je náhodný vektor o m složkách.

² Jde o totožnost středních hodnot stejnoolehých složek příslušných náhodných vektorů.

³ Podle 1) je $y_p = X_p \cdot \beta + \varepsilon_p$ a náhodnost vektoru y_p pramení z náhodnosti ε_p , zatímco v případě y_p^* vyplývá náhodnost tohoto vektoru z náhodnosti b (a tudíž z ε).

$$E(d_p) = E(y_p^* - y_p) = X_p E(b - \beta) - E\varepsilon_p = 0$$

Před odvozením kovarianční matice tohoto vektoru upravíme d_p do tvaru

$$\begin{aligned} d_p &= X_p \cdot b - X_p \beta - \varepsilon_p = -\varepsilon_p - X_p \beta + X_p (X'X)^{-1} X'y = \\ (5) \quad &= -\varepsilon_p - X_p \beta + X_p \beta + X_p (X'X)^{-1} X'\varepsilon = -\varepsilon_p + X_p (X'X)^{-1} X'\varepsilon \end{aligned}$$

Poznámka Vektor reziduí předpovídaných hodnot závisí tedy na náhodných složkách pozorovaného i náhodných složkách predikovaného období.

Za přijatých předpokladů **a) - d)** lze pro kovarianční matici *náhodného predikčního vektoru (5)* psát :

$$\begin{aligned} Cov(d_p) &= E \left[-\varepsilon_p + X_p (X'X)^{-1} X'\varepsilon \right] * \left[-\varepsilon_p + X_p (X'X)^{-1} X'\varepsilon \right] = \text{ }^4 \\ &\quad \text{[nekorelovanost } \varepsilon, \varepsilon_p \text{]} \\ &= E \left[\varepsilon_p \cdot \varepsilon_p' \right] + E \left[X_p (X'X)^{-1} X'\varepsilon \varepsilon' X (X'X)^{-1} X_p \right] = \\ &\quad \text{[} E\varepsilon_p \cdot \varepsilon_p' = \sigma^2 I_m \text{]} \quad \text{[} E\varepsilon \varepsilon' = \sigma^2 I_T \text{]} \\ &= \sigma^2 \cdot \left[I_m + X_p (X'X)^{-1} X_p' \right] \end{aligned}$$

Poznámka Speciální případ jediné predikované hodnoty tzn. případ $m=1$

$$Cov(d_p) = \sigma^2 \left[I + X_p (X'X)^{-1} X_p' \right]$$

V tomto případě jsou y_p a ε_p skalární hodnoty a matice X_p se redukuje na řádkový vektor x_p o k složkách.

Standardní chybu (směrodatnou odchylku) bodové předpovědi skutečné hodnoty vysvětlované proměnné dostaneme nahrazením rozptylu σ^2 jeho nestranným odhadem s^2 získaným například metodou OLS. Pak

$$s_{d_p} =_{OLS} s \cdot \sqrt{\left[I + X_p (X'X)^{-1} X_p' \right] \text{ }^5}$$

Lze ukázat, že vyhovuje-li model předpokladu normality, pak bodová predikční funkce založená na odhadu β metodou OLS má též normální rozdělení, konkrétně :

$$(6) \quad y_p^* \approx N \left[X_p \cdot \beta ; \sigma^2 X_p (X'X)^{-1} X_p' \right]$$

Protože, jak již jsme dokázali výše v **(3)**

$$E(y_p^*) = E(X_p \cdot b) = X_p E(b) = X_p \cdot \beta = E(y_p) \quad .$$

(3*) a dále

⁴ Bereme v úvahu nekorelovanost ε a ε_p .

⁵ Pokud $m=1$, pak x_p je vektor a veličina $x_p (X'X)^{-1} x_p'$ je skalár , pokud $m > 1$, pak se vezme diagonální prvek matice $X_p (X'X)^{-1} X_p'$ pořadím příslušný dané složce vektoru parametrů β .

$$Cov(y_p^*) = Cov(X_p \cdot b) = X_p \cdot Cov(b) \cdot X_p' = X_p \cdot \sigma^2 (X'X)^{-1} X_p' = \sigma^2 X_p \cdot (X'X)^{-1} X_p'$$

[Cov(b) = $\sigma^2 \cdot (X'X)^{-1}$]
[σ^2 je skalár]