

Dekompozice časové řady na složky

Náhodná složka

Je obsažena v každé reálné řadě zachycující v delším časovém úseku vývoj ekonomického ukazatele, mj. proto, že ekonomické prostředí je charakterizováno znatelnou mírou neurčitosti/stochastičnosti způsobené působením velkého množství i relativně slabých vlivů, které se v individuálním v časovém období projevují.

Občas se sice setkáme i s ekonomických časovými řadami, ve kterých není náhodná složka zastoupena, např. *časová řada daňových sazeb*, nebo *časová řada úrokových sazeb centrální banky (diskontní sazba PRIBOR)*, ale jde prakticky vždy o ukazatele, jejichž hodnoty jsou výsledkem arbitrárních rozhodnutí decizního subjektu, nikoliv působení exogenních ekonomických veličin či jiných vlivů. Drtivá většina „běžných“ ekonomických časových řad náhodnou složku obsahuje.

Aditivní dekompozice časové řady

$$Y_t = T_t + S_t + C_t + \varepsilon_t, \text{ kde}$$

T_t - představuje trendovou složku

S_t - představuje sezónní složku

C_t - představuje cyklickou složku

ε_t - představuje náhodnou (stochastickou) složku

Multiplikativní dekompozice časové řady

$$Y_t = T_t * S_t * C_t * \varepsilon_t, \text{ kde}$$

T_t - představuje trendovou složku

S_t - představuje sezónní složku

C_t - představuje cyklickou složku

ε_t - představuje náhodnou (stochastickou) složku

Analýze náhodné složky je v moderní analýze časových řad (nejméně stejně jako v ekonometrii) věnována větší pozornost, než kterékoliv z ostatních složek řady, přestože míra její významnosti na hodnotách časové řady může být jen několik málo procent. Je to dáno mj. tím, že v modelech časových řad, podobně jako v modelech ekonometrických jsou o náhodné složce vysloveny fundamentální předpoklady, které samy o sobě nebo s doplněním případných dalších umožňují konstruovat poměrně náročné matematicko-statistické postupy zasahující v řadě případů velmi hluboko do vnitřní struktury utváření jednotlivých hodnot řady.¹

¹ Zdaleka ne vždy však je praktický přínos z takto provedené analýzy úměrný míře složitosti těchto postupů a určitá část výzkumu v této oblasti je víceméně experimentováním s hypotézami, jejichž verifikace v realitě je možná jen ve velmi omezené míře.

Stacionarita časové řady

V širším slova smyslu lze říci, že *stacionarita časové řady* prezentuje vlastnost, kdy chování časové řady je „stochasticky“ ustálené.

Rozlišují se přitom dva základní typy stacionarity:

- **striktní stacionarita**, která znamená, že pravděpodobnostní chování příslušného **stochastického procesu**, který generuje hodnoty časové řady y_t , je **invariantní vůči posunům v čase**. tj. **pravděpodobnostní rozdělení náhodného vektoru** $(y_{t1}, y_{t2}, \dots, y_{tk}, \dots)$ je stejné jako rozdělení vektoru $(y_{t1+h}, y_{t2+h}, \dots, y_{tk+h})$ pro libovolné konečné h .

- **slabá (kovarianční) stacionarita**: je méně omezující než striktní stacionarita, neboť zde stačí, aby příslušný proces byl invariantní vůči posunům v čase pouze v rámci momentů rozdělení do druhého řádu, tj. pro libovolné platí pro každé s, t :

$$(1) \quad E(y_t) = \mu = const .$$

$$(2) \quad cov(y_s, y_t) = E(y_s - \mu)(y_t - \mu) = cov(y_{s+h}, y_{t+h})$$

pro libovolné h , speciálně tedy také pro $s = t$

$$(3) \quad var(y_t) = cov(y_t, y_t) = E(y_t^2) = \sigma_y^2 = const .$$

Střední hodnota a rozptyl stacionární řady jsou tedy konstantní (stejně pro všechny hodnoty) **v čase**. **Trend** (nekonstantní), **sezónnost**, **cykličnost** nebo **proměnný rozptyl** hodnot jsou tedy s vlastností stacionarity neslučitelné a pro její dosažení musí být přednostně z řady odstraněny. Také kovarianční struktura řady musí být v čase neměnná (např. charakter závislosti mezi 1 a 3.čtvrtletím musí zůstat stejný pro všechny členy čtvrtletní časové řady ve všech letech, která řada pokrývá) .

Poznámka1 Pokud u náhodného procesu existují konečné momenty aspoň do 2. řádu včetně, pak **ze striktní stacionarity evidentně vyplývá slabá**. Je-li navíc náhodný proces normální (tj. každý konečně n -rozměrný výběr z tohoto procesu má sdružené normální rozdělení), pak jsou oba typy stacionarity vzájemně ekvivalentní.

Ve většině úvah (a tedy všech následujících) budeme pracovat se **slabou stacionaritou** (kterou budeme označovat zkráceně jako **stacionarita**). Při popisu **Boxovy-Jenkinsovy metodologie** je vhodné začít právě modelováním stacionárních časových řad. I pojem autokovarianční a autokorelační funkce zavedeme (z tohoto důvodu) jen pro stacionární časové řady.

Autokovarianční a autokorelační funkce

Časové řady se často vyznačují silnou korelovaností v čase. Např. pokud časová řada tříměsíčního *PRIBORu* v daný den činila 3,45% p.a., pak v nejbližších dnech se spíše bude pohybovat v rozmezí od 3,40 do 3,50%, než někde kolem 6%. Běžným nástrojem pro kvantitativní popis tohoto jevu je *autokovarianční* nebo *autokorelační* funkce.

Autokovarianční funkce² pro zpoždění délky k (zkráceně **autokovariance pro zpoždění k**) se definuje jako

$$(4) \quad \gamma_k = \text{cov}(y_t, y_{t-k}) = E(y_t - \mu) \cdot (y_{t-k} - \mu), \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

Zřejmě máme pro $k = 0$: $\gamma_0 = \text{cov}(y_t, y_t) = E(y_t - \mu) \cdot (y_t - \mu) = \text{var } y_t = \sigma_y^2$

Analogicky, **Autokorelační funkce** pro zpoždění délky k (zkráceně **autokorelace pro zpoždění k**) (anglická zkratka **ACF**) se definuje jako

$$(5) \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_y^2} \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

Poznámka2

Označení autokorelační funkce pro výraz (5) je plně oprávněné, protože vzhledem k předpokladu stacionarity je zřejmé

$$(6) \quad \rho_k = \frac{\gamma_k}{\sigma_y^2} = \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{\text{var } y_t} \sqrt{\text{var } y_{t-k}}} = \frac{\text{cov}(y_t, y_{t-k})}{\sigma_y \cdot \sigma_{y-k}} = \text{corr}(y_t, y_{t-k})$$

Jak **autokovarianční** tak **autokorelační funkce** jsou funkce sudé (platí pro ně vztah $\gamma_k = \gamma_{-k}$, $\rho_k = \rho_{-k}$), takže se pro jejich popis stačí omezit na nezápornou hodnotu (argumentu označujícího zpoždění $k \geq 0$). Přitom zřejmě bude vždy platit

$$\rho_0 = 1 \quad \text{a} \quad |\rho_k| \leq 1, \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

Grafické zpodobnění ρ_k pro jednotlivá (diskrétní) k se nazývá **korelogram**.

Korelogram popisuje pomocí několika hodnot korelací krátkodobou dynamiku dané stacionární řady. Na jeho horizontální ose jsou uvedeny hodnoty (kladných) časových zpoždění, k nimž se v podobě sloupcového/úsečkového grafu vynášejí hodnoty autokorelační funkce pro danou hodnotu zpoždění.

Dlouhodobou dynamiku zachycuje obvykle trend (nebo cyklická složka). Bývá to právě několik (nemnoho) minulých hodnot y_{t-k} , pomocí jejichž korelací ρ_k s y_t vysvětluje model stacionární časové řady současnou hodnotu této řady - to je v kontrastu s klasickým modelem lineární regrese, kde daná vysvětlovaná proměnná y_t je vysvětlována pomocí jiných (vysvětlujících) proměnných x_{ij} a konvenční OLS-odhad b_j parametrů β_j zahrnuje korelaci mezi y_t a x_{ij} .

² Nejde zde o funkci ve smyslu slova matematické analýzy, protože tato „funkce“ je definována pouze pro celočíselné hodnoty časových zpoždění, nýbrž formálně zde jde o posloupnost.

Odhad parametrů autokovarianční a autokorelační funkce

Pro analyzovanou stacionární časovou řadu se používá

(7) *odhad střední hodnoty*

$$\bar{y} = \frac{1}{n} \cdot \sum_{t=1}^n y_t ,$$

(8) *odhad autokovarianční funkce*

$$c_k = \frac{1}{n} \cdot \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}) ,$$

(9) *odhad autokorelační funkce*

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k}^n (y_t - \bar{y})^2} ,$$

Aby měly tyto odhady praktický význam, požaduje se obvykle, aby $n > 50$ a navíc $k < n / 4$. V praxi se nicméně tento požadavek často opomíjí. Někdy se také ve vzorci (8) dělí číslem $n - k$ místo n . Tím se sice sníží vychýlenost odhadu, ale na druhé straně se zvětší *střední čtvercová chyba odhadu* definovaná jako

(10)

$$E(c_k - \gamma_k)^2$$

Tak či onak, $E(c_k)$ se při velkém n blíží skutečné hodnotě γ_k , stejně jako $E(r_k)$ se blíží ρ_k , takže se jedná o *asymptoticky nestranné odhady*.

Chování autokorelační funkce je v rámci *Boxovy-Jenkinsovy metodologie* důležitým ukazatelem, neboť napovídá, jaký typ modelu je vhodné pro danou řadu použít. Říkáme, že toto chování identifikuje příslušný model. Přitom je pro identifikaci důležité především určit hodnotu $k = k_0$, za kterou začíná být autokorelační funkce nulová (k_0 se pak označuje za *bod useknutí*). Současně je ale možné zjistit, že taková hodnota k_0 vůbec neexistuje.

Např. pro *model MA (1)* tvaru

(11A)

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} , \text{ kde } \varepsilon_t \text{ je bílý šum a } \theta_1 \text{ je parametr, platí}$$

(11B)

$$\rho_1 = \frac{\theta_1}{\theta_1^2 + 1} \quad \rho_k = 0 \text{ pro } k > 1$$

Takže zde je $k_0 = 1$. Přitom pro konkrétní analyzovanou řadu ale *teoretickou autokorelační funkci* ρ_k neznáme. Proto je nutné mít představu o tom, jak spolehlivě ji zastupuje *odhadnutá autokorelační funkce* r_k , kterou pro danou řadu můžeme podle vzorce (9) snadno spočítat. Zvlášť důležité je pak správně odpovědět na otázku, jak blízko nule musí být r_k , abychom s předem danou spolehlivostí mohli tvrdit, že pak $\rho_k = 0$. pro tento účel se často používá tzv.

Bartlettova aproximace:

je-li $\rho_k = 0$ pro $k > k_0$, pak za předpokladu (asymptotické) normality je

$$(12) \quad r_k \approx N \left(0, \sqrt{\frac{1}{n} \left(1 + 2 \cdot \sum_{j=1}^{k_0} r_j^2 \right)} \right) \quad \text{pro } k > k_0$$

Parciální autokorelační funkce a odhad jejích parametrů

Kromě autokorelační funkce ρ_k se používá také **parciální autokorelační funkce** značená jako ρ_{kk} (s anglickou zkratkou **PACF**). Hodnota ρ_{kk} je zde definována jako parciální autokorelační koeficient mezi y_t a y_{t-k} při pevných hodnotách $y_{t-k+1}, y_{t-k+2}, \dots, y_{t-1}$. Např. místo symbolu ρ_{22} bychom mohli užít obsažnějšího označení $\rho_{13.2}$. Zřejmě i zde platí $\rho_{00} = 1$ a $\rho_{11} = \rho_1$.

Vzhledem k definici parciální autokorelace ρ_{kk} je jejím přirozeným odhadem veličina r_{kk} odhadnutý parametr $\hat{\phi}_{kk}$ v modelu

$$(13) \quad y_t = \delta + \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \dots + \phi_{kk} y_{t-k} + \varepsilon_t, \quad \text{kde } \varepsilon_t \text{ je bílý šum.}$$

V reálných výpočtech se nicméně užívá rekurentní způsob výpočtu tohoto odhadu podle vzorců

$$(14) \quad r_{11} = r_1 \quad \text{a} \quad r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_j}, \quad \text{pro } k > 1, \text{ kde}$$

$$(14A) \quad r_{kj} = r_{k-1,j} - r_{kk} \cdot r_{k-1,k-j} \quad \text{pro } j = 1, 2, \dots, k-1$$

poznámka: výpočet probíhá rekurentně, každé $r_{kj}, j \neq k$ získáme pomocí r_{kk} nižších řádů, viz např. podle (14A):

$$k = 2, j = 1: \quad r_{21} = r_{11} - r_{22} \cdot r_{11} = (1 - r_{22}) \cdot r_{11}$$

$$k = 3, j = 1: \quad r_{31} = r_{21} - r_{33} \cdot r_{22} = (1 - r_{22}) \cdot r_{11} - r_{33} \cdot r_{22} = r_{11} - (1 + r_{33}) \cdot r_{22}$$

$$k = 3, j = 2: \quad r_{kj} = r_{22} - r_{33} \cdot r_{21} = r_{22} - r_{33} \cdot (1 - r_{22}) \cdot r_{11} = r_{22} - r_{11} \cdot r_{33} + r_{11} \cdot r_{22} \cdot r_{33}$$

Podobně jako autokorelační funkce může mít také **parciální autokorelační funkce bod useknutí**. (konkrétně pro autoregresní procesy), takže i ona je důležitým identifikačním nástrojem. Přitom se tentokrát užívá tzv.

Quenouilleova aproximace:

Je-li $\rho_{kk} = 0$ pro $k > k_0$, potom za předpokladu (asymptotické) normality je

$$(15) \quad r_{kk} \approx N \left(0, \sqrt{\frac{1}{n}} \right) \quad \text{pro } k > k_0.$$

ⁱ Typickými z nich jsou, že náhodné složky v různých obdobích jsou vzájemně nekorelované a mají stejné rozdělení (nebo aspoň stejné dva první momenty).