Periodicita v časové řadě, její popis a identifikace, exponenciální vyrovnávání

Statistika II

Jiří Neubauer

Katedra ekonometrie FVL UO Brno
kancelář 69a, tel. 973 442029
email:Jiri.Neubauer@unob.cz

Periodicita v časových řadách

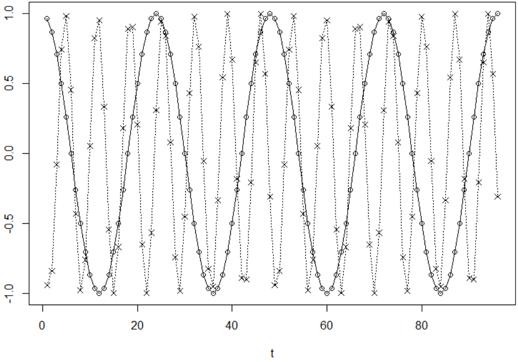
Některé časové řady obsahují periodickou složku. Pomocí vybraných nástrojů **spektrální analýzy** budeme tuto složku identifikovat. Mějme funkci periodickou funkci

$$R\cos(2\pi ft+\phi),$$

R je **amplituda**, f je **frekvence** a ϕ označuje **fázový posuv**. Tato funkce se opakuje každou časovou jednotku $T=\frac{1}{f}$ – **perioda**.

Periodicita v časových řadách

Graf zobrazuje dvě tyto funkce v diskrétním čase $t = 1, \ldots, 96$ s frekvencemi $4/96$ a $14/96$. Funkce s nižší frekvencí má nulový fázový posuv, ta s vyšší frekvencí má fázový posuv $0{,}6\pi$.

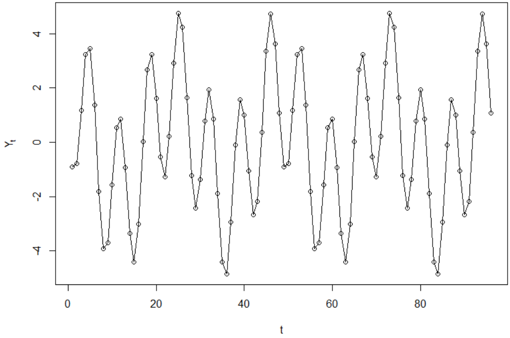


Periodicita v časových řadách

Vytvořme lineární kombinaci těchto funkcí

$$Y_t = 2\cos\left(2\pi t \frac{4}{96}\right) + 3\cos\left[2\pi\left(t\frac{14}{96} + 0{,}3\right)\right]. \qquad (1)$$

Periodicita je nyní „skrytá“.



Periodicita v časových řadách

Platí

$$R\cos(2\pi ft + \phi) = A\cos(2\pi ft) + B\sin(2\pi ft),$$

kde

$$R = \sqrt{A^2 + B^2}, \quad \phi = \arctan\left(-\frac{B}{A}\right)$$

a obráceně

$$A = R\cos\phi, \quad B = -R\sin\phi.$$

Pro pevně danou hodnotu frekvence f lze použít $\cos(2\pi ft)$ a $\sin(2\pi ft)$ jako prediktory a odhadnout A a B pomocí metody nejmenších čtverců.

Periodicita v časových řadách

Obecnou kombinaci m kosinových funkcí s libovolnými amplitudami a frekvencemi lze zapsat ve tvaru[1]

$$Y_t = A_0 + \sum_{j=1}^{m} \left[A_j \cos(2\pi f_j t) + B_j (2 \sin f_j t) \right].$$

Metodu nejmenších čtverců lze použít pro odhady A_j a B_j, pokud mají frekvence speciální tvar, regrese jsou jednoduché. Předpokládejme že n je liché, $n = 2k + 1$. Frekvence $1/n, 2/n, \ldots, k/n$ se nazývají **fourierovské frekvence**. Prediktory tvořené funkcemi sinus a kosinus v těchto frekvencích jsou ortogonální, dostáváme odhady

$$\widehat{A}_0 = \overline{Y_t}$$

$$\widehat{A}_j = \frac{2}{n} \sum_{t=1}^{n} Y_t \cos\left(\frac{2\pi tj}{n}\right) \quad \text{a} \quad \widehat{B}_j = \frac{2}{n} \sum_{t=1}^{n} Y_t \sin\left(\frac{2\pi tj}{n}\right)$$

[1] A_0 je koeficient kosinové funkce pro nulovou frekvenci, B_0 je koeficient kosinové funkce pro nulovou frekvenci, je tedy nulový a vzorci se neobjevuje.

Periodicita v časových řadách

Je-li n sudé, $n = 2k$, předchozí rovnice platí pro $j = 1, 2 \ldots, k-1$, ale pro $j = k$ dostáváme

$$\widehat{A}_k = \frac{1}{n}\sum_{t=1}^{n}(-1)^t Y_t, \quad \text{a} \quad \widehat{B}_k = 0. \tag{2}$$

Periodogram

Pro liché $n = 2k + 1$ je **periodogram** I pro frekvenci $f = j/n, j = 1, 2, \dots, k$ definován

$$I\left(\frac{j}{n}\right) = \frac{n}{2}\left(\widehat{A}_j^2 + \widehat{B}_j^2\right).$$

Pro sudé $n = 2k$ získáme hodnoty periodogramu pro $j = 1, 2, \dots, k - 1$ podle předchozího vztahu, pro $j = k$, tedy frekvenci $f = k/n = 1/2$ je

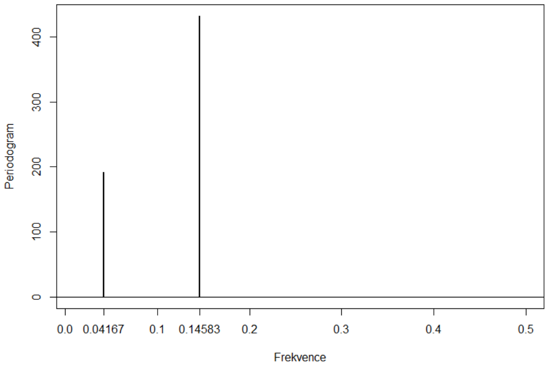
$$I\left(\frac{1}{2}\right) = n\widehat{A}_k^2,$$

viz výraz (2).

Pozn. Pro dlouhé časové řady se pro výpočet používá FFT (fast Fourier transform).

Periodogram

Graf zobrazuje periodogram lineární kombinace funkcí (1)



Periodogram

Rozšíříme nyní definici periodogramu na všechny frekvence z intervalu $0 \leq f \leq 1/2$

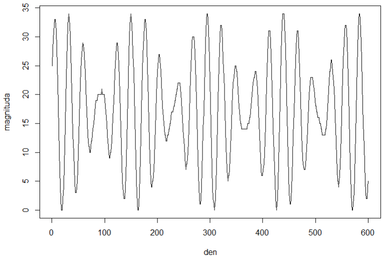
$$I(f) = \frac{n}{2}\left(\widehat{A}_f^2 + \widehat{B}_f^2\right),$$

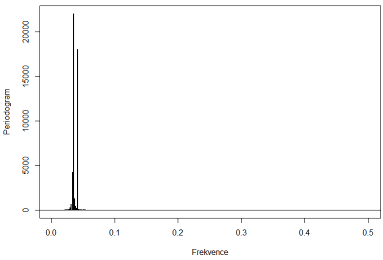
kde

$$\widehat{A}_f = \frac{2}{n}\sum_{t=1}^{n} Y_t \cos(2\pi t f) \quad \text{a} \quad \widehat{B}_f = \frac{2}{n}\sum_{t=1}^{n} Y_t \sin(2\pi t f).$$

Periodogram

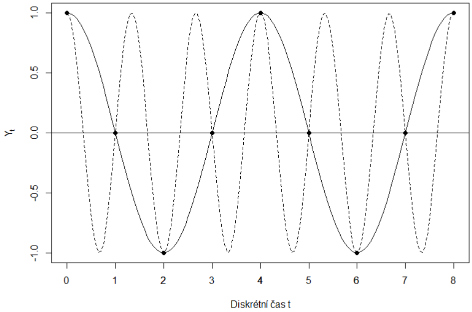
Půlnoční magnitudy (jas) jisté hvězdy v 600 po sobě jdoucích dnech



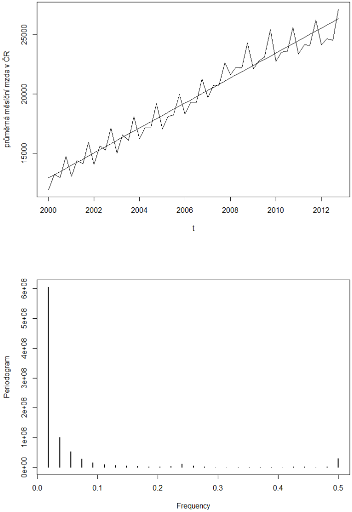


Periodogram

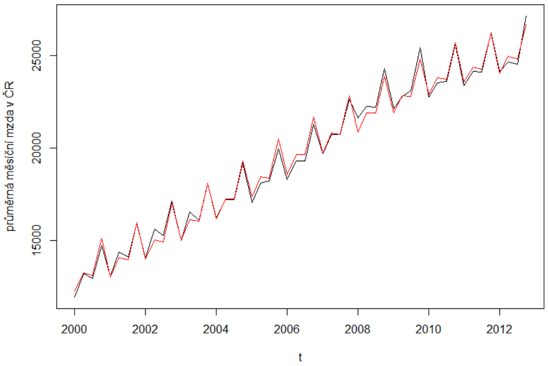
Proč frekvence omezovat na interval $0 \leq f \leq 1/2$? V grafu jsou zobrazeny dvě kosinové funkce, jedna s frekvencí $f = 1/4$ a druhá s frekvencí $f = 3/4$ (čárkovaná čára). Měříme-li hodnoty těchto funkcí pouze v časech $t = 0, 1, 2, \ldots,$ dostáváme identické hodnoty. V diskrétním čase nemůže od sebe tyto dvě funkce rozlišit – **aliasing** frekvencí. **Nyquistova frekvence** – $f = 1/2$.



Periodogram - mzda v ČR



Periodogram - mzda v ČR



Exponenciální vyrovnávání

Metody exponenciálního vyrovnávání určují předpovědi časové řady pomocí minulých pozorování. Tato predikce (předpověď) je obvykle určena jako vážený součet těchto pozorování. Mezi základní metody patří jednoduché exponenciální vyrovnávání, které nezahrnuje ani trend ani sezónní složku. Předpověď pro čas $t+h$ je založená na úrovni proměnné $\hat{a}$ v čase t

$$\hat{Y}_t(h) = \hat{a}_t,$$

kterou lze rekurzivně odhadnout jako vážený průměr pozorované a predikované hodnoty Y_t

$$\begin{aligned} \hat{a}_t &= \alpha Y_t + (1-\alpha)\hat{Y}_{t-1} = \\ &= \alpha Y_t + (1-\alpha)\hat{a}_{t-1}, \end{aligned} \tag{3}$$

kde $0 < \alpha < 1$ je vyhlazovací parametr.

Exponenciální vyrovnávání

Za počáteční hodnotu $\hat{a}_1$ se obvykle bere hodnota Y_1. Postupným dosazováním do vztahu (3) dostáváme

$$\hat{Y}_t(h) = \hat{a}_t = \alpha Y_t + (1-\alpha)\hat{a}_{t-1} = \alpha Y_t + (1-\alpha)(\alpha Y_{t-1} + (1-\alpha)\hat{a}_{t-2}) =$$
$$= \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + (1-\alpha)^2 \hat{a}_{t-2} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} +$$
$$+ (1-\alpha)^2(\alpha Y_{t-2} + (1-\alpha)\hat{a}_{t-3}) =$$
$$= \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \dots.$$

Odtud je vidět, že předpověď $\hat{Y}_t(h)$ je tvořena váženým součtem hodnot $Y_{t-i}, i = 0, 1, 2, \dots$ s váhami $\alpha(1-\alpha)^i$. Volba vyhlazovacího parametru α ovlivňuje spočítanou předpověď. V programu R je hodnota parametru α zvolena tak, aby se minimalizovala čtvercová chyba predikce (součet čtverců odchylek naměřených a predikovaných hodnot).

Exponenciální vyrovnávání

Metodu jednoduchého exponenciální vyrovnávání lze rozšířit o trendovou složku, předpovědi mají potom tvar

$$\hat{Y}_t(h) = \hat{a}_t + h\hat{b}_t,$$

kde

$$\begin{aligned} \hat{a}_t &= \alpha Y_t + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}), \\ \hat{b}_t &= \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1} \end{aligned}$$

Uvedenou metodu lze použít v případě, kdy v krátkých úsecích lze trendovou složku považovat za lineární. Podobně jako u jednoduchého exponenciálního vyrovnávání, obsahuje uvedená metoda neznámé parametry α a β. Jejich hodnoty jsou v programu R nalezeny minimalizací čtvercové chyby predikce.

Exponenciální vyrovnávání

Model lze ještě dále rozšířit o sezónní složku, hovoří se potom o tzv. Holt-Wintersovu vyrovnávání. Obsahuje tři parametry: α pro úroveň, β pro trend a γ pro sezónní složku (funkce `HoltWinters`). Aditivní Holt-Wintersovo vyrovnání s periodou sezónní složky p má tvar

$$\hat{Y}_t(h) = \hat{a}_t + h\hat{b}_t + \hat{s}_{(t-p+1+(h-1) \mod p)},$$

kde $\hat{a}_t$, $\hat{b}_t$ a $\hat{s}_t$ jsou dány vztahy

$$\begin{aligned}
\hat{a}_t &= \alpha(Y_t - \hat{s}_{t-p}) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}),\\
\hat{b}_t &= \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1},\\
\hat{s}_t &= \gamma(Y_t - \hat{a}_t) + (1-\gamma)\hat{s}_{t-p}.
\end{aligned}$$

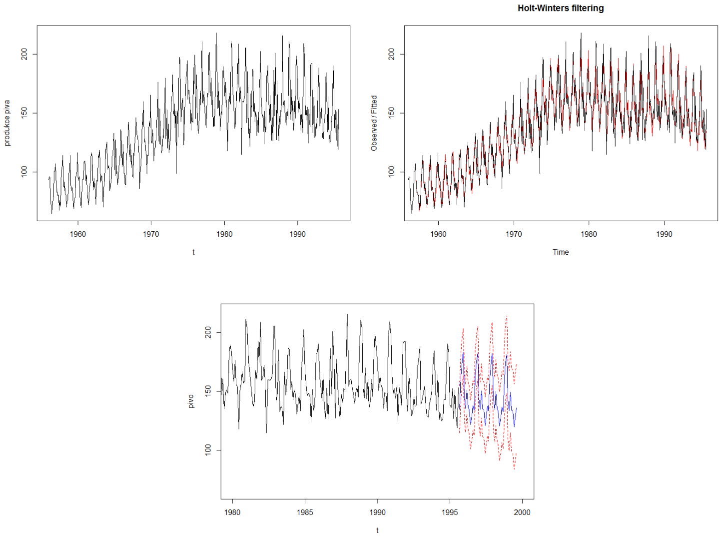
Multiplikativní Holt-Wintersovo vyrovnávání má tvar

$$\hat{Y}_t(h) = (\hat{a}_t + h\hat{b}_t)\hat{s}_{(t-p+1+(h-1) \mod p)},$$

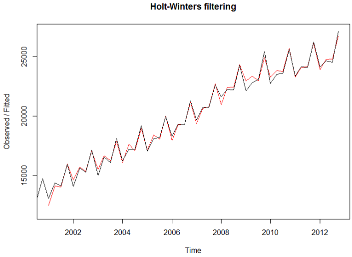
kde $\hat{a}_t$, $\hat{b}_t$ a $\hat{s}_t$ jsou dány vztahy

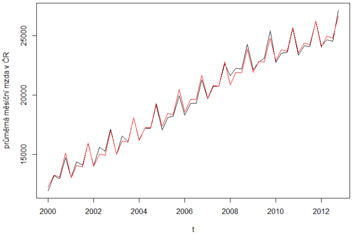
$$\begin{aligned}
\hat{a}_t &= \alpha(Y_t - \hat{s}_{t-p}) + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}),\\
\hat{b}_t &= \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1},\\
\hat{s}_t &= \gamma(Y_t/\hat{a}_t) + (1-\gamma)\hat{s}_{t-p}.
\end{aligned}$$

Exponenciální vyrovnávání – měsíční produkce piva v Austrálii



Exponenciální vyrovnávání – mzda v ČR





Obrázek vlevo ukazuje výsledek exponenciálního vyrovnávání, obrázek vpravo potom fit pomocí periodických funkcí – viz periodogram.