

# Časové řady

Mgr. Jiří Mazurek, Ph.D.

# Analýza časových řad

- Analýza časových řad představuje v současnosti velmi důležitou součást ekonometrie, neboť umožňuje popisovat systémy, které mění v čase svůj charakter.
- Cílem analýzy časových řad je především porozumět mechanismu, který vygeneroval hodnoty dané časové řady, neboť to umožňuje alespoň do jisté míry „ovládat“ fungování systému, o jehož chování vypovídají naměřené hodnoty.
- Umožňuje to provádět předpovědi budoucího chování takového systému. Systém, který řadu vytvořil, je popisován matematickým modelem.

# Časová řada

- Časová řada je posloupnost prostorově a věcně srovnatelných číselných údajů uspořádaných v čase od minulosti přes přítomnost do budoucnosti.
- Zde nás budou zajímat zejména časové řady ekonomických veličin, speciálně tržeb, neboli tzv. **ekonomické časové řady**.
- Rozdělení:
  - **okamžikové** časové řady
  - **intervalové** časové řady

# Předpoklady

- V časové řadě se obvykle předpokládá, že:
  - hlavním faktorem změny je čas (označuje se  $t$ ),
  - údaje jsou uvedeny za ekvidistantní, tj. stejně dlouhé časové intervaly.
- Vývoj časové řady se popisuje matematickým modelem. Hlavním cílem konstrukce takového modelu je jeho využití k predikci budoucího vývoje řady.
- **Prognózování** představuje odhad budoucí velikosti závislé proměnné.
- Rozdělení prognóz:
  - bodové prognózy
  - intervalové prognózy.

# Dekompoziční metody časových řad

- Předpokládá se, že model časové řady může obsahovat až 4 složky, které vyjadřují různé druhy pohybu analyzovaného ukazatele:
  - trendovou složku (trend)  $T_t$
  - sezónní složku  $S_t$
  - cyklickou složku  $C_t$
  - náhodnou složku  $\varepsilon_t$
- Trendová, sezónní a cyklická složka tvoří společně **deterministickou složku.**

# Trendova složka

- **Trendová složka** vyjadřuje základní směřování hodnot časové řady (růst, pokles a jejich eventuální zesílení nebo tlumení). Tato složka vyjadřuje systematický a dlouhodobější vliv faktorů, které působí jedním směrem. Trend může být buďto rostoucí nebo klesající. Nepřevažuje-li ani růst ani pokles, jedná se o časovou řadu bez trendu.

# Periodická složka

- **Sezónní a cyklická složka**, souhrnně nazývané **periodická složka**, zachycují pravidelné kolísání hodnot časové řady.
- **Sezonní složka** vyjadřuje pravidelné výkyvy hodnot časové řady, k nimž dochází během roku. Tyto výkyvy se pravidelně opakují. Důležitým rysem sezonní složky, nebo se také říká sezónnosti, je skutečnost, že časová prodleva mezi výkyvy není delší než jeden rok.
- **Cyklická složka** reprezentuje vliv faktorů, které způsobují dlouhodobější výkyvy hodnot řady. Říká se také, že jde o výkyvy kolem trendu, přičemž časová prodleva mezi těmito výkyvy je na rozdíl od sezónnosti delší než jeden rok. Protože intenzita výkyvů i jejich pravidelnost se často mění, cyklickou složku je v obtížné detekovat stejně tak jako její příčiny.

# Aditivní model

- Zpravidla se uvažuje, že složky časové řady jsou v aditivním vztahu, takže model časové řady potom můžeme zapsat ve tvaru:

$$Y_t = T_t + S_t + C_t + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

- V tomto případě se hovoří o **aditivním modelu** časové řady. V ekonomických časových řadách se nejčastěji setkáme se dvěma speciálními případy modelu:
  - 1. s případem, kdy se v řadě nevyskytuje periodická složka:

$$Y_t = T_t + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

- 2. s případem, kdy se v řadě nevyskytuje cyklická složka, tedy tzv. **časovou řadu se sezónní složkou**

$$y_t = T_t + S_t + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$



# Multiplikativní model

- Vedle aditivního modelu existuje také **multiplikativní model** vycházející z předpokladu, že vzájemný vztah jednotlivých složek modelu je dán pronásobením:

$$y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

# Trend

- Trend se popisuje nejčastěji lineární funkcí, polynomem druhého stupně, exponenciální funkcí, modifikovanou exponenciální funkcí nebo logistickou, případně Gompertzovou křivkou.
- V případě lineární funkce a polynomu druhého stupně jde o regresní funkce lineární z hlediska parametrů, takže pro odhad neznámých parametrů můžeme v jejich případě aplikovat obyčejnou metodu nejmenších čtverců.
- V případě ostatních křivek je situace složitější, protože tyto funkce nejsou lineární z hlediska parametrů, takže pro odhad jejich parametrů se musí postupovat jinak.

# Trendy

- Kromě lineárního trendu se vyskytují také:
  - Polynomický trend:  $Y = \beta_0 + \beta_1 \cdot x + \beta_2 x^2 + \dots + \beta_k x^k$ .
  - Logaritmický trend:  $Y = \beta_0 + \beta_1 \ln(x)$ .
  - Mocninný trend:  $Y = \beta_0 \cdot x^{\beta_1}$ .
  - Exponenciální trend:  $Y = \beta_0 \beta_1^{x_1} \beta_2^{x_2} \dots \beta_k^{x_k}$ , speciálně jednoduchý exponenciální trend:  $Y = \beta_0 \beta_1^x$ .

# Syntetické modely trendu časových řad

- **Nejsou** zadány explicitně vzorcem
- **Jsou** zadány hodnotami nové časové řady (syntetického trendu)
- **Klouzavé průměry** – časové řady posouvaných průměrů (mediánů) několika hodnot „okolo“  $t$
- **Exponenciální vyrovnání** – časové řady posouvaných vážených průměrů hodnot „před“  $t$  (váhy exponenciálně ubývají)

# Prosté klouzavé průměry

- Pokud chceme použít klouzavé průměry, musíme především zvolit tzv. **délku klouzavé části**  $a$  dále tzv. **řád klouzavého průměru**. Řád je dán stupněm polynomu, kterým se části řady vyrovnávají.
- V případě prostého klouzavého průměru používáme k vyrovnávání lineární funkci, takže pracujeme s řádem jedna.
- Délka klouzavého průměru se obvykle volí jako liché číslo obecně zapsané ve tvaru  $2k + 1$ , kde  $k$  je celé kladné číslo. Každá část řady, která je vyrovnávána, má svůj střed.

# Klouzavé průměry

**Prosté klouzavé průměry** (lichá délka „kolem“  $t$ ):

$$\bar{y}_t = \frac{1}{2p+1} \sum_{t=-p}^p y_{t+i} = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}}{2p+1}$$

o délce  $m = 2p+1$ , kde  $t = p+1, p+2, \dots, n-p$ .

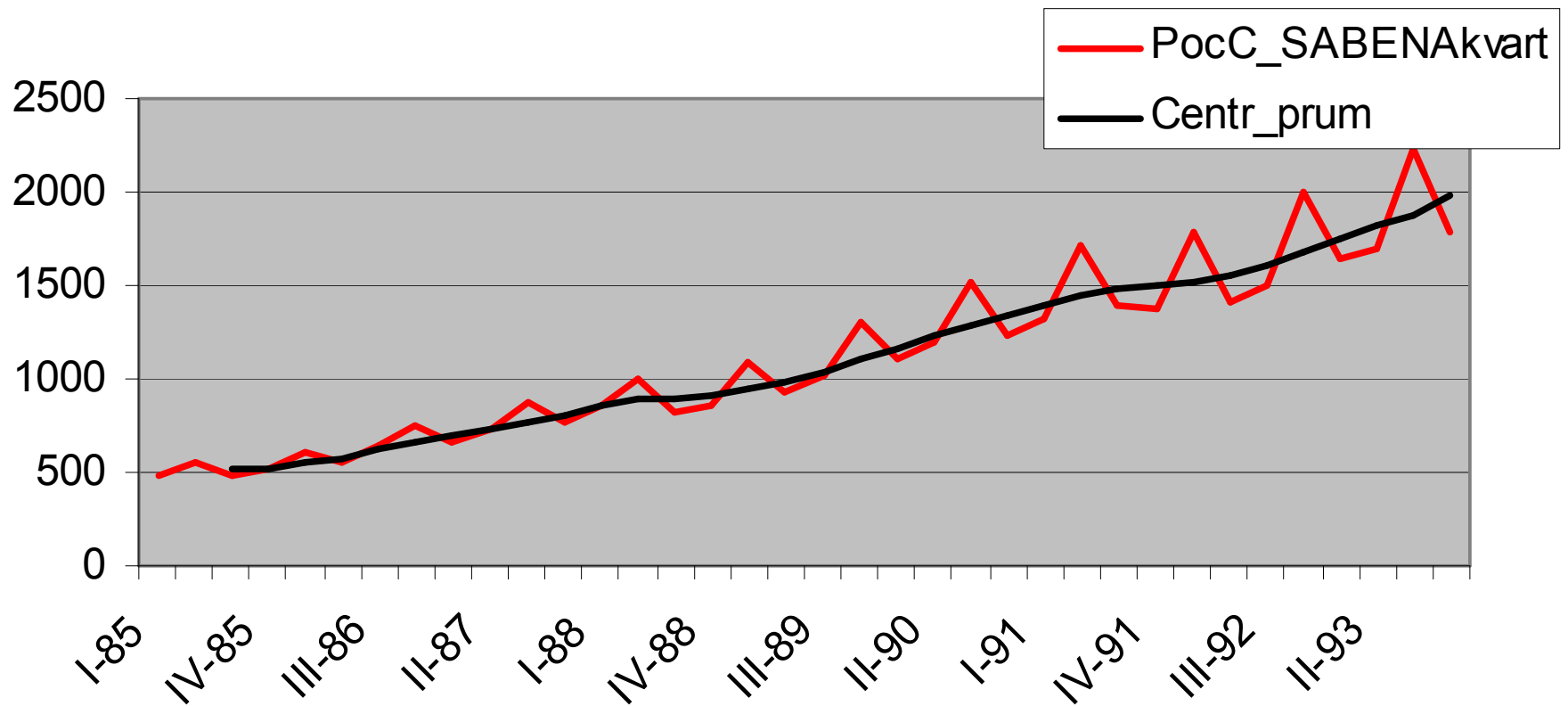
**Centrované klouzavé průměry** (sudá délka):

$$\bar{y}_t = \frac{\frac{y_{t-p} + y_{t-p+1}}{2} + \dots + \frac{y_{t+p-1} + y_{t+p}}{2}}{p}, t = p+1, \dots$$

o délce  $m = 2p$

# Příklad: centrováný 4-členný klouzavý průměr

## Čtvrtletní počty cestujících SABENA 4-členné centrováné průměry



# VLASTNOSTI NÁHODNÉ SLOŽKY MODELU A JEJICH OVĚŘENÍ

1. Střední hodnota  $\varepsilon_i$  je nula, tj.  $E(\varepsilon_i) = 0$  pro každé  $i$ .
2. Rozptyl  $\varepsilon_i$  je konstantní, nezávislý na  $i$ , tj.  $Var(\varepsilon_i) = \sigma^2$  pro každé  $i$ .
3. Veličiny  $\varepsilon_i, \varepsilon_j$  jsou nekorelované, tj.  $Cov(\varepsilon_i, \varepsilon_j) = 0$  pro  $i \neq j$ .
4. Veličiny  $\varepsilon_i$  mají normální rozdělení, tj.  $\varepsilon_i \sim N(0, \sigma^2)$  pro každé  $i$ .

Platí-li bod 2, hovoříme o **homoskedasticitě** (v opačném případě o heteroskedasticitě).

Platí-li bod 3, mluvíme o **nezkorelovanosti** náhodných složek modelu.



# Testování vlastností pro rezidua

- Uvedené podmínky by měly být ověřeny vhodnou statistickou metodou.
- Podmínka 1 se neověřuje a je brána za danou.
- Jsou-li splněny všechny podmínky, potom odhady získané metodou nejmenších čtverců budou nejlepší v rámci všech nestranných odhadů.
- Jsou-li splněny jen podmínky 1-3, budou odhady parametrů nejlepší „pouze“ v rámci tzv. lineárních nestranných odhadů.
- Tedy, i když podmínka 4 splněna není, pořád nám popsane postupy poskytují odhady parametrů, které jsou „rozumně“ kvalitní.
- Pokud jde o podmínku 2, existuje např. statistický test Goldfeld-Quandtův, který je konkrétnější, pokud jde o formulaci podoby případné testované heteroskedasticity, a také existují testy obecnější, pokud jde o tuto formulaci. Mezi obecnější testy patří např. Whiteův test. Problém heteroskedasticity je ale typický pro průřezovaná data, nikoliv pro modely časových řad, pro které je typické nedodržení podmínky 3.
- My: testování podmínky pro autokorelaci reziduí.

# Durbin-Watsonův test

- K ověřování autokorelace se využívá zejména Durbinův-Watsonův test.
- Test zkoumá platnost nulové hypotézy, že model není zatížen autokorelací, proti alternativní hypotéze, že v modelu je autokorelace ve tvaru AR(1).
- Nejprve se najdou odhady parametrů původního regresního modelu časové řady metodou nejmenších čtverců a ze získaných vyrovnaných hodnot se vypočtou reziduální odchylky. Na základě těchto reziduí se pak počítá testové kritérium

$$T = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2},$$

# Durbin-Watsonův test II

- Pro toto kritérium jsou určeny speciální statistické tabulky.
- V těchto tabulkách se pro daný počet pozorování, hladinu významnosti a počet parametrů modelu bez absolutního členu najde dolní hodnota  $d_L$  a horní hodnota  $d_H$ .
- Dále je třeba vypočítat odhad párové korelace mezi reziduí regresního modelu:

$$r = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2}.$$

- Pro  $r > 0$  když  $T$  je větší než  $d_H$ , nulová hypotéza o absenci autokorelace se přijímá, zatímco je-li kritérium menší než  $d_L$ , hypotéza se zamítá.
- Pro  $r < 0$  se vypočte statistika  $T^* = 4 - T$  pro kterou platí to, co v předchozím případě.
- Pokud se kterékoliv z testovacích kritérií dostane mezi hodnoty  $d_L$  a  $d_H$ , nelze na základě testu rozhodnout o platnosti či neplatnosti  $H_0$

# Durbinův-Watsonův test III

## Durbin-Watsonův test

Autokorelace prvního řádu se nejčastěji testuje pomocí Durbin-Watsonova testu, který se označuje jako  $d$  nebo  $DW$  test.  $DW$  test bývá součástí běžných ekonometrických softwarů (např. Gretl).

$DW$  test není použitelný pro regresní modely bez úrovně konstanty a modely se zpožděnou závisle proměnnou mezi vysvětlujícími proměnnými.

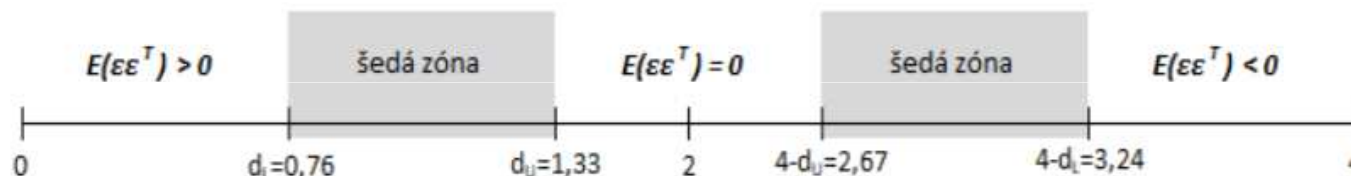
$$d = \frac{\sum(e_t - e_{t-1})^2}{\sum e_t^2}, \quad d \in (0,4)$$

kde  $e_t$  je hodnota rezidua v čase  $t$ ,  $e_{t-1}$  je hodnota rezidua a v čase  $t-1$

$DW$  statistika má symetrické rozdělení se střední hodnotou  $E(d) = 2$ . Hodnoty v její blízkosti představují sériovou nezávislost náhodné složky. Hodnoty blízké 0 představují pozitivní autokorelaci náhodné složky a hodnoty blízké 4 představují negativní autokorelaci.

K přesnému posouzení výsledků testovací statistiky  $d$  se používají tabulkové hodnoty, které vymezují intervaly pozitivní a negativní autokorelace, interval nulové autokorelace a intervaly šedé zóny (neprůkaznosti). Dolní  $d_L$  a horní  $d_U$  meze jsou koncipovány pro počet pozorování  $n$  a počet nezávisle proměnných  $k$ .

Například pro  $n=8$  a  $k=1$  jsou  $d_L = 0,76$  a  $d_U = 1,33$ .



# Prognózování pomocí časových řad

- Prognózování se nazývá predikování, předpovídání, předvídání, extrapolace, apod.
- Mezi prognostickými metodami hrají významnou roli statistické prognostické metody. Do této skupiny patří také metody používající při konstrukci prognóz extrapolaci časových řad využívající regresní analýzy.
- Podstata extrapolačních metod spočívá ve studiu minulosti prognózovaného jevu a v přenosu zákonitostí vývoje z minulosti a přítomnosti do budoucnosti.
- U procesů, které jsou v čase stabilní, lze tento princip s úspěchem použít. Naopak v případě, kdy během prognózovaného období probíhají podstatné kvalitativní změny, je použití extrapolačních modelů problematické.

# Bodový odhad

- Uvažujme model časové řady  $Y_t = T_t + \varepsilon_t$ ,  $t = 1, 2, \dots, n$ , kde  $T_t$  odráží lineární nebo kvadratický trend a  $n$  je časový okamžik přítomnosti.
- Bodový odhad  $\tilde{Y}_{n+h}$  neznámé veličiny časové řady v čase  $n + h$ , kde  $h$  je zadaný **horizont bodové prognózy**, lze stanovit takto:
  - $\tilde{Y}_{n+h} = T_{n+h}$
- Zde  $T_{n+h}$  je trendová funkce vyčíslená v čase  $n + h$ .
- Bodová předpověď umožňuje pomocí jednoho čísla odhadnout hodnotu předpovídané veličiny.
- Spočívá jednoduše v tom, že do odhadnuté regresní funkce/do odhadnutého trendu dosadím budoucí časový okamžik, který mne zajímá.

# Intervalový odhad

- Kromě bodové predikce konstruujeme také intervaly spolehlivosti pro  $\hat{Y}_{n+h}$
- Intervalová prognóza vytvořená v čase  $n$  na období posunuté o  $i$  časových jednotek dopředu je definována jako oboustranný interval spolehlivosti - viz dále.

# Intervalový odhad – lineární trend

- V případě lineárního trendu má 95% interval spolehlivosti tvar:

$$[ \tilde{Y}_{n+i} - t_{n-2}(0,05) s \sqrt{Q_n(i)}, \tilde{Y}_{n+i} + t_{n-2}(0,05) s \sqrt{Q_n(i)} ],$$

- kde

$$s = \sqrt{\frac{\sum_{t=1}^n Y_t^2 - \sum_{t=1}^n \hat{T}_t^2}{n-2}}$$

$$Q_n(i) = 1 + \frac{1}{n} + \frac{(n+i-\bar{t})^2}{\sum_{t=1}^n t^2 - n\bar{t}^2}, \quad \bar{t} = (n+1)/2$$



# Intervalový odhad – kvadratický trend

- V případě kvadratického trendu má 95% interval spolehlivosti tvar

$$[ \tilde{Y}_{n+i} - t_{n-3}(0,05) s \sqrt{Q_n(i)}, \tilde{Y}_{n+i} + t_{n-3}(0,05) s \sqrt{Q_n(i)} ],$$

- kde

$$s = \sqrt{\frac{\sum_{t=1}^n Y_t^2 - \sum_{t=1}^n \hat{T}_t^2}{n-3}}$$

$$Q_n(i) = 1 + [1, n+i, (n+i)^2] \cdot (X^T \cdot X)^{-1} \cdot [1, n+i, (n+i)^2]^T, X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \cdot & \cdot & \cdot \\ 1 & n & n^2 \end{pmatrix}$$

# Příklad 1

Tabulka 27 obsahuje fiktivní údaje o vývoji měsíčních výdajů domácností na potraviny v Moravskoslezském kraji (v milionech Kč). Data byla získána v období leden 2000 ( $t = 1$ ) až březen 2001 ( $t = 15$ ).

**Tabulka 27: měsíční výdaje domácností  $Y_t$**

$Y_t$	141	145	142	147	146	154	150	158	157	165	164	170	167	174	175
$t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

- a) Určete lineární trendovou složku, b) proveďte Durbinův-Watsonův test na autokorelaci, c) proveďte predikci výdajů na následující 3 měsíce.

# Příklad 2

5.5. Časovou řadu uvedenou v tabulce proložte kvadratickým trendem. Použijte k tomu metodu nejmenších čtverců.

<b>t</b>	1	2	3	4	5	6	7	8	9	10
<b>Y<sub>t</sub></b>	1,2	6,3	14,3	37,1	76,5	125	274	349	499	578

<b>t</b>	11	12	13	14	15	16	17	18	19	20
<b>Y<sub>t</sub></b>	711	859	987	1114	1135	1349	1506	1680	1721	1890

Přidáme si i lineární trend a exponenciální trend, vše v Excelu.

# Příklad 3

5.6. Časovou řadu uvedenou v následující tabulce vyrovnejte prostými klouzavými průměry délky pět.

<b>t</b>	1	2	3	4	5	6	7
<b>Y<sub>t</sub></b>	18,683	15,236	20,552	20,988	30,598	23,22	38,375

<b>T</b>	8	9	10	11	12	13	14	15
<b>Y<sub>t</sub></b>	43,698	47,813	61,403	62,002	68,386	63,904	68,247	67,818

# Příklad 4 v Gretlu

- Analyzujte časovou řadu a ověřte model s lineárním trendem a sezónností (cykličností), a proveďte predikci, viz připravený soubor v Excelu.

Děkuji za pozornost