Časové řady

Doc. 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 ε_t
- Trendová, sezónní a cyklická složka tvoří společně **deterministickou složku**.

Trendová 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: $Y = \beta_0 + \beta_1.x + \beta_2 x^2 + \ldots + \beta_k x^k$.
- Logaritmický trend: $Y = \beta_0 + \beta_1 \ln(x)$.
- Mocninný trend: $Y = \beta_0 .x^{\beta_1}$.
- Exponenciální trend: $Y = \beta_0 \beta_1^{x_1} \beta_2^{x_2} \ldots \beta_k^{x_k}$, 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 , kde je celé kladné číslo. Každá část řady, která je vyrovnávána, má svůj střed.

Klouzavé průměry

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

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

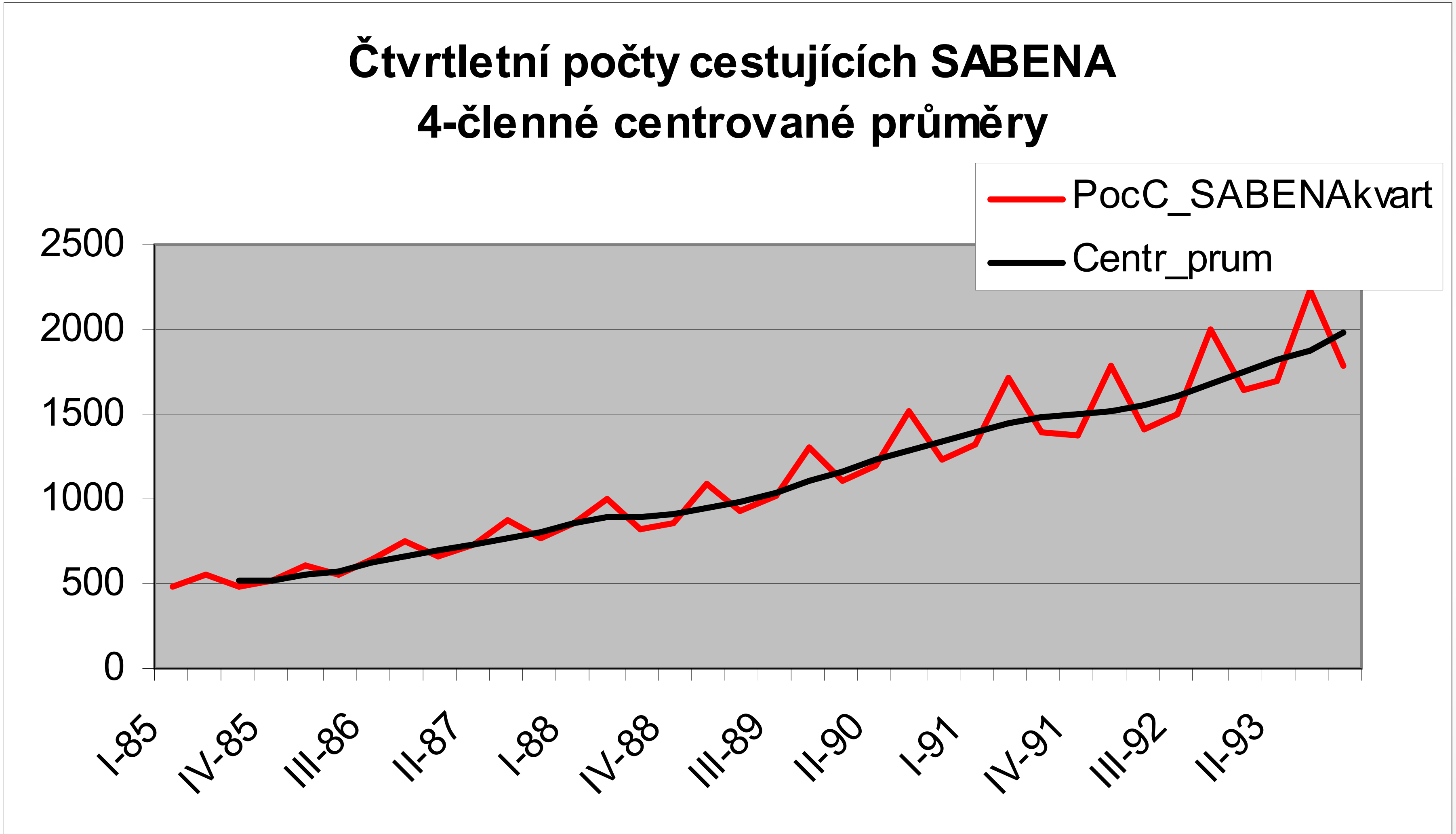
o délce $m = 2p+1$, kde $t = p+1, p+2, \ldots, n-p.$

Centrované klouzavé průměry (sudá délka):

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

o délce $m = 2p$

Příklad: centrovaný 4-členný klouzavý průměr



VLASTNOSTI NÁHODNÉ SLOŽKY MODELU A JEJICH OVĚŘENÍ

1. Střední hodnota ε_i je nula, tj. $E(\varepsilon_i) = 0$ pro každé i.
2. Rozptyl ε_i je konstantní, nezávislý na i, tj. $Var(\varepsilon_i) = \sigma^2$ pro každé i.
3. Veličiny ε_i, ε_j jsou nekorelované, tj. $Cov(\varepsilon_i, \varepsilon_j) = 0$ pro $i \neq j$.
4. Veličiny ε_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 popsané 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 úrovňové konstanty a modely se zpožděnou závisle proměnnou mezi vysvětlujícími proměnnými.

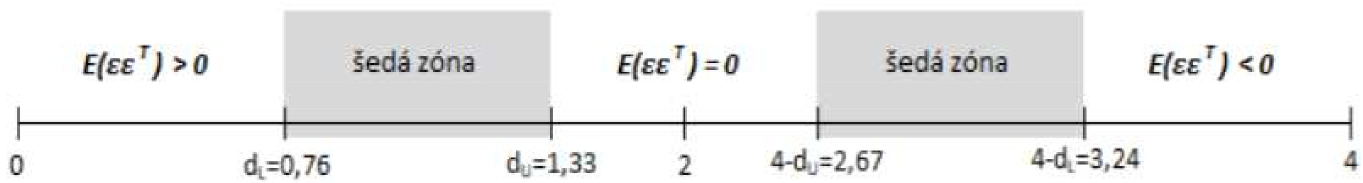
$$d = \frac{\sum(e_t - e_{t-1})^2}{\sum e_t^2}, \quad d \in \langle 0,4 \rangle$$

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,…, *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.

Příklad 1 - kriminalita

Data z Veřejné databáze ČSU

Kriminalita - trestné činy [1]

Počet registrovaných trestných činů — **Území:** Česká republika

	Kriminalita celkem	Obecná kriminalita	Hospodářská kriminalita	Loupeže	Vloupání do bytů a rodinných domů	Znásilnění	Vraždy
2005	344 060	277 504	43 882	5 550	10 361	596	186
2006	336 446	263 371	39 473	4 783	9 603	530	231
2007	357 391	268 301	37 981	4 856	9 163	637	196
2008	343 799	257 763	32 474	4 641	9 111	529	202
2009	332 829	249 975	29 774	4 687	9 616	480	181
2010	313 387	249 038	28 371	4 019	10 091	586	173
2011	317 177	252 957	28 216	3 881	9 568	675	173
2012	304 528	242 449	27 633	3 416	9 718	669	188
2013	325 366	260 465	30 376	3 051	11 117	589	182
2014	288 660	222 494	30 731	2 547	8 877	669	160
2015	247 628	184 357	30 616	2 022	6 895	598	155
2016	218 162	160 614	28 306	1 646	6 099	649	136
2017	202 303	150 167	26 294	1 585	5 465	598	146
2018	192 405	141 581	24 837	1 406	4 724	651	116
2019	199 221	145 829	24 589	1 439	4 892	683	143
2020	165 525	121 981	18 528	1 248	4 060	639	130
2021	153 233	117 349	12 510	1 230	3 784	773	105
2022	181 991	144 527	13 637	1 422	4 273	880	150
2023	181 417	144 973	12 765	1 439	3 588	917	159

Kód: KRI05/7

Příklad 1 - kriminalita

- Z dat ČSÚ o kriminalitě použijeme časovou řadu „Kriminalita celkem“.
- Určíme (v Excelu) následující charakteristiky:

1) Lineární regresní model: R^2, statistická významnost.
2) Rezidua
3) Durbin-Watsonův test
4) Predikce hodnot do budoucna
5) Vyrovnání časové řady pomocí klouzavého průměru o délce 3.

Rešení si ukážeme v Excelovském souboru „Příklad_1_Přesnáška_5.xls“.

Děkuji za pozornost