

2. Analýza trendové složky

Rychlý náhled do problematiky kapitoly

Klasická analýza ekonomických časových řad vychází z předpokladu, že časovou řadu je možné rozložit na čtyři složky: trendovou, cyklickou, sezónní a nesystematickou (náhodnou) složku. V této kapitole se budeme věnovat analýze trendové složky. Tento přístup analýzy časových řad patří mezi klasické analytické postupy, které jsou založeny na regresní analýze.

Budete umět

- Charakterizovat jednotlivé složky časové řady
- Odhadnout koeficienty lineárního trendu
- Zvolit vhodnou trendovou funkci
- Posoudit kvalitu sestaveného modelu časové řady

2.1 Přístupy k modelování časových řad

Tradičním výchozím principem modelování časových řad je jednorozměrný model ve tvaru

$$y_t = f(t, \varepsilon_t), \quad (2.1)$$

kde y_t je hodnota modelovaného ukazatele v čase t , $t = 1, 2, \dots, n$, ε_t je hodnota náhodné složky (poruchy) v čase t .

K modelu (2.1) přistupujeme trojím způsobem:

- a) pomocí klasického (formálního) modelu – dekompozice časové řady,
- b) pomocí Box – Jenkinsovy metodologie,
- c) pomocí spektrální analýzy.

a) Dekompozice časové řady

- v tomto modelu jde o popis forem pohybu a ne o poznání věcných příčin dynamiky časové řady
- model vychází z dekompozice časové řady na 4 složky
- hledáme nástroje, který vysvětlí „systematické“ chování sledovaného procesu
-

Časovou řadu je možné rozložit na čtyři složky:

- ◆ **Trendová** T_t - odráží tendenci vývoje zkoumaného jevu za dlouhé období,
 - je výsledkem faktorů, které dlouhodobě působí ve stejném směru,
 - např. technologie výroby, demografické podmínky, podmínky trhu, počet dovezených osobních aut do ČR po roce 1990
- ◆ **Sezónní** S_t - vyjadřuje periodické kolísání v časové řadě, které má systematický charakter,
 - toto kolísání se během 1 kalendářního roku a každý rok opakuje,

- vyskytuje se pouze u časových řad krátkodobých (čtvrtletních, měsíčních, týdenních),
 - může měnit každý rok měnit svůj charakter,
 - zachycuje střídání ročních období (nákup dovolených, vánoce)
 - pracovní cyklus (výplata mezd a nákupy v maloobchodě vždy v určitou dobu)
- ◆ **Cyklická C_t**
- je „pohyb“ okolo trendu, ve kterém se střídají fáze růstu s fázemi poklesu,
 - cykly mají nepravidelný charakter,
 - cykly se odehrávají v obdobích delších než jeden rok,
 - někdy může být zaměněna se složkou trendovou,
 - cyklické pohyby v ekonomických časových řadách mohou mít příčiny ekonomického i neekonomického charakteru,
 - na rozdíl od makroekonomie, kdy se rozumí porucha dynamické rovnováhy ekonomiky s jednotlivými fázemi, chápe statistika cyklus jako dlouhodobé kolísání s neznámou periodou, která může mít i jiné příčiny než ekonomický cyklus (cykly demografické, inovační, plánovací)
- ◆ **Reziduální ε_t**
- má nesystematický charakter, nelze ji popsat žádnou funkcí času
 - je tvořena náhodnými pohyby v časové řadě, chybami v měřeních a jinými nesystematickými vlivy, nepostižitelné příčiny,
 - je náhodnou složkou a její vlastnosti budeme prověřovat testy

Dekompozice časové řady :

A) **ADITIVNÍ** – časová řada je tvořena součtem jednotlivých složek

$$y_t = T_t + C_t + S_t + \varepsilon_t.$$

Jednotlivé složky jsou uvažovány se svých absolutních hodnotách a jsou v měrných jednotkách původní časové řady.

B) **MULTIPLIKATIVNÍ** – časová řada je tvořena součinem jednotlivých složek

$$y_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t$$

V absolutní hodnotě je pouze trendová složka, ostatní složky jsou vyjádřeny relativně.

Proč se časová řady rozkládá?

- zkoumáním oddělených složek lze odhalit některé zákonitosti vývoje dané řady
- lze identifikovat některé vnější vlivy určující průběh časové řady
- je možno lépe porovnat průběh několika časových řad
- umožňuje přesnější konstrukce předpovědí
- je to prostředek umožňující sezónní očištění časových řad

b) Box – Jenkinsovy metodologie

- považuje za základní prvek konstrukce modelu časové řady náhodnou složku,
- důraz klade na korelační analýzu závislých pozorování uspořádaných do tvaru časové řady,
- patří zde ARIMA modely (p, d, q),
- model klouzavých průměrů MA(1), $y_t = \varepsilon_t + 0,49\varepsilon_{t-1}$
- autoregresní modely AR (1), $y_t = 0,42y_{t-1} + \varepsilon_t$
- časová řada musí obsahovat aspoň 30 pozorování

c) Spektrální analýza

- časovou řadu považujeme za „směs“ sinusovek a kosinusovek s různými amplitudami a frekvencemi,
- lze provést explicitní popis periodického chování časové řady,
- chceme najít významné složky periodicity, které se podílí na vlastnosti zkoumaného procesu,
- hlavním faktorem není časová proměnná, ale faktor frekvenční

Vedle jednorozměrného modelu typu (2.1) se můžeme setkat i s modely, které jsou založeny na předpokladu, že vývoj analyzovaného ukazatele není ovlivňován pouze časovým faktorem, ale i řadou jiných ukazatelů. Tyto ukazatele, kterými se snažíme vývoj analyzovaného ukazatele vysvětlit, nazýváme příčinné nebo faktorové.

Takovýto model lze zapsat ve formě

$$y_t = f(t, x_1, x_2, \dots, x_n, \varepsilon_t), \quad (2.2)$$

kde x_1, x_2, \dots, x_n jsou ukazatele ovlivňující analyzovaný ukazatel.

Modely tohoto typu se obecně nazývají vícerozměrné modely. Změny faktorové proměnné x_i v modelu (2.2) na proměnnou y se nemusí projevit pouze v časovém okamžiku t . Omezíme-li se pouze na jednu faktorovou proměnnou x , dostaneme tzv. Koyckův model rozložených zpoždění, který má obecný tvar $y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots$, kde koeficienty β_i jsou neznámé parametry, které vyjadřují vliv faktorové proměnné x z období $t, t-1, t-2, \dots, t-z$ na analyzovanou veličinu y v čase t .

2.2 Trendová analýza

Časová řada vzniká tak, že v každém čase daná náhodná veličina stochastického procesu nabude právě jednu hodnotu v závislosti na jistých faktorech. *Stochastickým procesem* je v čase uspořádaná řada náhodných veličin.

Např. máme časovou řadu produkce výroby. Určitá hodnota této časové řady je jednou z množiny hodnot. Tato hodnota je dána takovými faktory (např. technologie výroby, řízení výroby, management firmy, kvalifikace pracovníků, mezilidské vztahy na pracovišti,...) Jsou to faktory *systematické a nesystematické*. Systematické faktory způsobují existenci trendu (sezónní a cyklické složky), nesystematické způsobují existenci reziduální složky.

Předpokládáme následující dekompozici časové řady: $y_t = Y_t + \varepsilon_t, \quad t = 1, \dots, n,$

kde systematická složka Y_t představuje trend T_t , který lze vyjádřit matematickou funkcí časové proměnné t ,

ε_t je reziduální složka typu *bílého šumu* - jde o časovou řadu generovanou *stochastickým procesem*, $\varepsilon_t = N(0, \sigma)$.

Opakování regresní analýzy

$$y_t = Y_t + \varepsilon_t, \quad t = 1, \dots, n$$

y_t ... naměřená hodnota, Y_t ... teoretická hodnota regresní funkce, ε_t ... odchylka,

$\beta_0, \beta_1, \dots, \beta_p$... neznámé parametry regresní funkce

$Y = f(t; \beta_0, \beta_1, \dots, \beta_p)$... regresní funkce - hledáme typ funkce (trendové funkce)

- odhadujeme parametry funkce

- odhady parametrů označujeme b_0, b_1, \dots, b_p

$\hat{Y} = f(t; b_0, b_1, \dots, b_p)$... odhad teoretické hodnoty regresní funkce

$y_t = \hat{Y}_t + e_t$, kde e_t je reziduum (odhad náhodné složky)

Cílem je, aby e_t měla charakter bílého šumu, tzn. aby v odhadnutých reziduích bylo minimum části systematické.

Metoda nejmenších čtverců

$$Q = \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (y_t - Y_t)^2 = \min .$$

- vypočteme parciální derivace podle parametrů β_j ,
- rovnice položíme = 0,
- dostaneme $p+1$ *normálních rovnic*,
- v těchto rovnicích β_j nahradíme b_j , $j = 1, 2, \dots, n$

Metoda nejmenších čtverců je tedy založena minimalizací chyby, tj. odchylky teoretické hodnoty od skutečné hodnoty.

Trendové funkce

Konstantní trend

$$Y_t = \beta_0, \quad t = 1, \dots, n$$

Lineární trend

$$Y_t = \beta_0 + \beta_1 t, \quad t = 1, \dots, n$$

Kvadratický trend

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, \dots, n$$

Exponenciální trend

$$Y_t = \alpha \beta^t, \quad t = 1, \dots, n \quad \alpha > 0, \beta > 0$$

Modifikovaný exponenciální trend

$$Y_t = \gamma + \alpha \beta^t, \quad t = 1, \dots, n \quad \alpha < 0, \quad 0 < \beta < 1, \quad \gamma > 0$$

(funkce není lineární v parametrech a není ji možno linearizovat, nelze použít k odhadu parametrů metodu nejmenších čtverců)

Logistický trend

$$Y_t = \frac{\gamma}{1 + \alpha \beta^t}, \quad t = 1, \dots, n \quad \beta > 0, \quad \gamma > 0$$

(funkce je charakteristická nezápornou asymptotou a jedním inflexním bodem, pro odhad parametrů nelze použít metody nejmenších čtverců)

Více o některých trendových funkcích

Lineární trend

$$Y_t = \beta_0 + \beta_1 t, \quad t = 1, \dots, n.$$

Lineární trend odhadneme rovnicí $\hat{Y}_t = b_0 + b_1 t$, kde \hat{Y}_t, b_0, b_1 jsou odhady teoretických hodnot Y_t, β_0, β_1 . Skutečnou (naměřenou, získanou) hodnotu označme y_t .

Pro tuto hodnotu platí vztah $y_t = Y_t + \varepsilon_t \Rightarrow \varepsilon_t = y_t - Y_t$.

Veličina ε_t se nazývá reziduum. Odhad rezidua ε_t označme e_t , kde $e_t = y_t - \hat{Y}_t$. Naším cílem je minimalizovat chybu (reziduum). Tato metoda se nazývá **metoda nejmenších čtverců**. Označme $Q = \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (y_t - Y_t)^2$.

Funkci Q chceme minimalizovat. Dosaďme $Y_t = \beta_0 + \beta_1 t$ a dostáváme vztah:

$Q = \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)^2$. Tuto funkci parciální derivujeme podle proměnných β_0, β_1 a první parciální derivace funkce Q položíme rovny nule. Dostáváme soustavu **normálních rovnic**:

$$\frac{\partial Q}{\partial \beta_0} = 2 \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)(-1) = 0,$$

$$\frac{\partial Q}{\partial \beta_1} = 2 \sum_{t=1}^n (y_t - \beta_0 - \beta_1 t)(-t) = 0.$$

Po úpravě dostaneme následující rovnice:

$$n\beta_0 + \beta_1 \sum_{t=1}^n t = \sum_{t=1}^n y_t,$$

$$\beta_0 \sum_{t=1}^n t + \beta_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n y_t \cdot t.$$

V soustavě normálních rovnic nahradíme teoretické hodnoty parametrů β_0, β_1 jejich odhady b_0, b_1 :

$$nb_0 + b_1 \sum_{t=1}^n t = \sum_{t=1}^n y_t,$$

$$b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n y_t \cdot t.$$

Uvedenou soustavu lze řešit Cramerovým pravidlem a dostáváme

$$b_0 = \bar{y} - b_1 \bar{t}, \quad b_1 = \frac{\sum_{t=1}^n t y_t - \bar{t} \sum_{t=1}^n y_t}{\sum_{t=1}^n t^2 - n \bar{t}^2}.$$

Pro ruční výpočet je snadnější zavést substituci:

$$t' = t - \bar{t}, \text{ je-li } n \text{ liché; } t' = 2(t - \bar{t}), \text{ je-li } n \text{ sudé; kde } \bar{t} = \frac{1}{n} \sum_{t=1}^n t.$$

Pak platí $\sum (t')^k = 0, \quad k = 1, 3, 5, \dots$

Tabulka 2.1 Transformovaná časová proměnná při sudém a

Rok	2001	2002	2003	2004	2005	2006	2007	2008
t	1	2	3	4	5	6	7	8
t'	-7	-5	-3	-1	1	3	5	7

Tabulka 2.2 Transformovaná časová proměnná při lichém a

Rok	2001	2002	2003	2004	2005	2006	2007
t	1	2	3	4	5	6	7
t'	-3	-2	-1	0	1	2	3

Po zavedení výše uvedené substituce dostáváme následující vztahy pro výpočet odhadů teoretických hodnot parametrů:

$$\sum_{t=1}^n y_t = n b_0 \quad \Rightarrow \quad b_0 = \frac{\sum_{t=1}^n y_t}{n} = \bar{y},$$

$$\sum_{t=1}^n y_t \cdot t' = \sum (t')^2 b_1 \quad \Rightarrow \quad b_1 = \frac{\sum t' y_t}{\sum (t')^2}.$$

Kvadratický trend

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, \dots, n.$$

Podobně jako u lineárního trendu odhadneme teoretické hodnoty parametrů $\beta_0, \beta_1, \beta_2$, které označíme b_0, b_1, b_2 , metodou nejmenších čtverců. Pro odhady teoretických hodnot parametrů platí následující vztahy:

$$b_0 = \frac{\sum y_t \sum t'^4 - \sum t'^2 \sum y_t t'^2}{n \sum t'^4 - (\sum t'^2)^2}, \quad b_1 = \frac{\sum y_t t'}{\sum t'^2}, \quad b_2 = \frac{n \sum y_t t'^2 - \sum y_t \sum t'^2}{n \sum t'^4 - (\sum t'^2)^2}.$$

Exponenciální trend

$$Y_t = \alpha \beta^t, \quad t = 1, \dots, n \quad \alpha > 0, \beta > 0.$$

Danou rovnicí budeme nejprve logaritmovat a po zavedení substituce dostaneme lineární rovnici, jejíž parametry odhadneme metodou nejmenších čtverců. Postupujeme tedy následovně: $\ln Y_t = \ln(\alpha \beta^t) \Rightarrow \ln Y_t = \ln \alpha + t \ln \beta$.

Zavedeme substituce: $\ln Y_t = Y'_t$, $\ln \alpha = \alpha'$, $\ln \beta = \beta'$ a pro časovou proměnnou t zavedeme substituci, která je definována u lineárního trendu. Dostáváme lineární trend, který je definován vztahem $Y'_t = \alpha' + t \beta'$. Odhady parametrů α', β' označme a', b' . Pro odhadné parametry a, b v exponenciálním trendu platí $a = e^{a'}$, $b = e^{b'}$.

Logistický trend

$$Y_t = \frac{\gamma}{1 + \alpha \beta^t}, \quad t = 1, \dots, n \quad \beta > 0, \gamma > 0.$$

Logistická trendová funkce byla původně odvozena jako křivka vyjadřující biologický růst populací za podmínek omezených zdrojů. V ekonomické oblasti se tato křivka začala používat v modelech poptávky po předmětech dlouhodobé spotřeby a s úspěchem se také používá např. při modelování vývoje, výroby a prodeje některých druhů výrobků. Patří mezi trendové funkce s kladnou horní asymptotou a jedním inflexním bodem. Pole typického průběhu se této skupině křivek říká S-křivky. Každá S-křivka vymezuje na časové ose pět základních vývojově odlišných fází cyklu. Cyklem ze budeme obecně rozumět časové období od prosazení nových sil (technologií, výrobků,...) až o jejich zániku, kdy jsou vystřídány silami novými na kvalitativně vyšší rovní. Jednotlivé fáze lze charakterizovat následujícím způsobem:

1. fáze – období, kdy se začínají formovat nové progresivní síly. Jejich prosazování je ještě v té době brzděno, původními ne zcela překonanými silami.

2. fáze – období, kdy se nové progresivní síly začínají plně prosazovat a rozhodující měrou ovlivňovat další vývoj. Tyto síly působí jako akcelerátory, takže tempo vývoje se v této fázi značně urychluje.

3.fáze – období, kdy nové progresivní síly zcela ovládly další vývoj, ale už se objevují opoziční síly, tlumící jejich účinek. Vývoj v této fázi nabývá lineárního charakteru.

4.fáze – období, kdy vzniklé opoziční síly nabývají postupnou převahu nad dosavadními silami, které pozbyly svou progresivnost, čímž se vývojové tendence podstatně zpomalují.

5.fáze – poslední fáze cyklu, kdy opoziční síly nabyly rozhodující převahu a zcela utlumily vývoj dosavadních sil. V této fázi se vývoj zastavuje až o okamžiku zformování a prosazení dalších kvalitativně progresivních sil.

V praxi se s podobným vývojem můžeme setkat např. v managementu inovací. Zde logický trend může vhodně simulovat důležitá stadia ekonomické životnosti inovovaného či nového výrobku. Ten přechází postupně od počátečních vývojových fází (uvádění na trh, reklamní fáze,...) přes růst trhu a stadium zralosti až po období poklesu nebo stagnace tohoto produktu na trhu.

Na rozdíl o přecházejících typů trendových funkcí, které jsou v podstatě definovány jednoznačně je logistická trendová funkce vyjadřována v různých tvarech. Nejčastěji uváděné tvary:

$$Y_t = \frac{k}{1 + a_0 e^{-a_1 t}}, \quad Y_t = \frac{k}{1 + e^{a_0 + a_1 t}}, \quad Y_t = \frac{k}{1 + e^{-a_1 t}}, \quad Y_t = \frac{k}{1 + a_0 a_1^t}.$$

Ať použijeme jakýkoliv tvar, vždy si logická funkce zachová svůj charakteristický průběh ve tvaru písmen S, protože jednotlivé formy zápisu se od sebe liší pouze různými vzájemnými transformace parametrů.

Věnujme pozornost odhadu parametrů logistické trendové funkce, přičemž budeme vycházet z naposledy uvedeného tvaru, kde $a_0 > 0$, $a_1 > 0$. Odhad parametrů lze provést metodou vybraných bodů. Uvažujme tři vybrané body t , $t + m$, $t + 2m$ (obvykle $t = 0$) a položíme v nich empirické hodnoty rovny teoretickým.

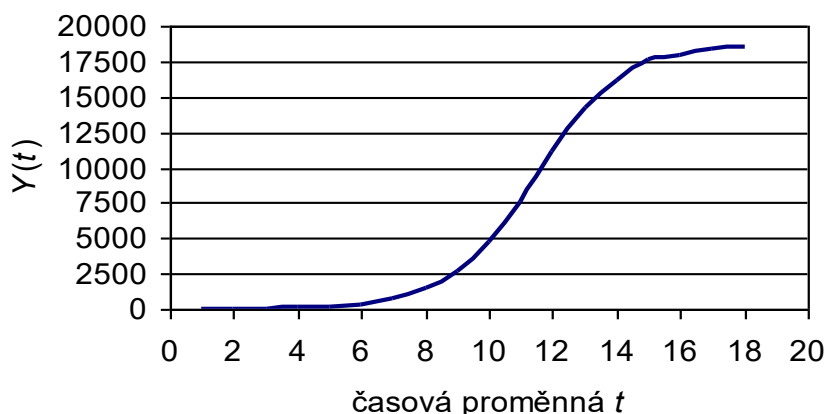
Vypočítáme pomocné veličiny:

$$S_1 = \frac{1}{y_{t+2m}} - \frac{1}{y_{t+m}}, \quad S_2 = \frac{1}{y_{t+m}} - \frac{1}{y_t}, \quad S_3 = \frac{y_{t+m}}{y_t}.$$

Odhady parametrů potom získáme ze vzorců:

$$\hat{a}_1 = \sqrt[m]{\frac{S_1}{S_2}}, \quad \hat{a}_0 = \frac{S_3 - 1}{(1 - S_3 \hat{a}_1^m) \hat{a}_1^t}, \quad \hat{k} = y_t (1 + \hat{a}_0 \hat{a}_1^t).$$

Logistický trend



2.3 Volba a ověřování vhodnosti trendové funkce

- volba trendové funkce na základě grafu zkoumané časové řady (subjektivní ovlivnění)
- jednoduchým a účinným prostředkem je *analýza diferencí a koeficientů růstu* dané čř
- **analýza diferencí** je vhodná pro výběr trendové funkce polynomiálního typu, protože první diference určitého polynomu snižuje jeho řád o jeden stupeň:
 - v případě *konstanty* je první diference: $T_t - T_{t-1} = \beta_0 - \beta_0 = 0$
 - v případě *přímky* je první diference: $T_t - T_{t-1} = \beta_0 + \beta_1 t - \beta_0 - \beta_1(t-1) = \beta_1$
 - v případě *paraboly* je první diference: $T_t - T_{t-1} = (\beta_1 - \beta_2) + \beta_2 t$ (přímka)
- **koeficient růstu** exponenciální funkce je $\frac{T_t}{T_{t-1}} = \frac{\alpha \beta^t}{\alpha \beta^{t-1}} = \beta$.

Pravidlo výběru trendové funkce:

- jestliže jsou *první diference* přibližně nulové, volíme konstantní trend,
- jestliže jsou *druhé diference* přibližně nulové, volíme lineární trend,
- jestliže jsou *třetí diference* přibližně nulové, volíme parabolický trend,
- jestliže jsou *koeficienty růstu* přibližně konstantní, volíme exponenciální funkci.

Použití tohoto přístupu je problematické, pokud diference vykazují velkou variabilitu. Tento přístup je nepoužitelný při rozhodování mezi složitějšími typy trendových křivek.

Pokud výše uvedené způsoby nevedou k jednoznačnému výsledku, vybereme několik trendových funkcí a jejich vhodnost ověříme až po odhadnutí jejich parametrů.

Příklad 1. Pro časovou řadu uvedenou v následující tabulce stanovíme na základy analýzy absolutních diferencí vhodný typ trendové funkce.

Rok	y_t	$\Delta^{(1)}y_t$	$\Delta^{(2)}y_t$	$\Delta^{(3)}y_t$	$\Delta^{(4)}y_t$
2000	3				
2001	4	1			
2002	7	3	2		
2003	14	7	4	2	
2004	27	13	6	2	0
2005	48	21	8	2	0

2006	79	31	10	2	0
2007	122	43	12	2	0

Řešení: Z průběhu prvních a druhých absolutních diferencí je zřejmé, že vykazují výrazný trend. Třetí diference jsou již konstantní, takže trend dané časové řady nejlépe vystihneme pomocí kubického polynomického trendu.

Pro ověřování vhodnosti trendové funkce se používají **dva druhy kritérií:**

- a) interpolační kritéria, b) extrapoláční kritéria
(Následné hodnocení přesnosti předpovědí)

a) Interpolační kritéria - jsou založena na zkoumání vztahu skutečných hodnot časové řady a tzv. interpolovaných, tj. minulých a současných hodnot odhadnutých na základě dané trendové funkce.

Pro posouzení tohoto vztahu lze použít následující charakteristiky:

A) I) střední chyba (Mean Error)

$$ME = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{Y}_t)$$

II) střední čtvercová chyba (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{Y}_t)^2$$

III) střední absolutní chyba (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{Y}_t|$$

IV) střední absolutní chyba procentuální (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{Y}_t|}{y_t} \cdot 100$$

V) střední chyba procentuální (Mean Percentage Error)

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(y_t - \hat{Y}_t)}{y_t} \cdot 100$$

Zvolená trendová funkce je tím lepší, čím tyto charakteristiky nabudou menších hodnot.

Příklad 2. V tabulce jsou uvedeny údaje o počtu prodaných CD nosičů hudebním vydavatelstvím v tis.ks v letech 1999 – 2007. Odhadněme trend časové řady pomocí lineární, kvadratické a exponenciální trendové funkce. Určeme chyby odhadu (charakteristiky uvedeny v bodu A) a rozhodněme, která z trendových funkcí je pro danou časovou řadu nejlepší.

Rok	1999	2000	2001	2002	2003	2004	2005	2006	2007
y_t	3	10	15	21	35	42	58	81	110

Řešení: Následující tabulka zachycuje přehled chyb.

Model	ME	MSE	MAE	MAPE	MPE
$-20,66 + 12,46t$	0,000	93,214	8,296	62,578	37,386
$8,02 - 3,18t + 1,56t^2$	0,000	9,403	2,848	20,622	-8,503
$3,61 \cdot 1,49^t$	-2,011	80,817	6,033	21,824	-3,621

Řešení: Z tabulky vidíme, že k nejlepšímu výsledku vede použití paraboly. Parabola má v určitém oboru hodnot průběh velmi podobný s exponenciálou.

B) *Durbinův – Watsonův test* se používá pro testování hypotézy, že rezidua nejsou autokorelovaná (mají charakter bílého šumu). Alternativní hypotéza je

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad 0 < |\rho| < 1,$$

tj. rezidua jsou autokorelovaná (obsahují systematickou složku). Testové kritérium má

tvar:
$$DW = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}, \quad \text{kde} \quad \varepsilon_t = y_t - \hat{Y}_t.$$

Statistika DW nabývá hodnot od 0 do 4.

Výsledky D-W testu :

$$\begin{aligned} 4 - d_L < DW < 4 &\Rightarrow H_0 \text{ se zamítá - autokorelace} \\ 4 - d_U < DW < 4 - d_L &\Rightarrow \text{Test nic neříká} \\ 2 < DW < 4 - d_U &\Rightarrow H_0 \text{ se přijímá - autokorelace není} \\ d_U < DW < 2 &\Rightarrow H_0 \text{ se přijímá - autokorelace není} \\ d_L < DW < d_U &\Rightarrow \text{Test nic neříká} \\ 0 < DW < d_L &\Rightarrow H_0 \text{ se zamítá - autokorelace} \end{aligned}$$

Kritické hodnoty d_L a d_U jsou tabelovány.

Příklad 3. Testujme Durbin-Watsonovým testem hypotézu o nezávislosti reziduí. Hodnoty reziduí jsou vypočteny v následující tabulce.

t	1	2	3	4	5	6	7	8	9	10	11	12
e_t	4	7	9	6	-11	-24	-6	-17	-15	2	19	26

Řešení: Tabulku doplníme o další řádky, abychom mohli vypočítat hodnotu Durbin-Watsonova testového kritéria.

t	1	2	3	4	5	6	7	8	9	10	11	12	Součet
e_t	4	7	9	6	-11	-24	-6	-17	-15	2	19	26	
e_t^2	16	49	81	36	121	576	36	289	225	4	361	676	2470

$(e_t - e_{t-1})^2$		9	4	9	289	169	324	121	4	289	289	49	1556
---------------------	--	---	---	---	-----	-----	-----	-----	---	-----	-----	----	------

Dosazením dostáváme $DW = \frac{1556}{2470} = 0,63$.

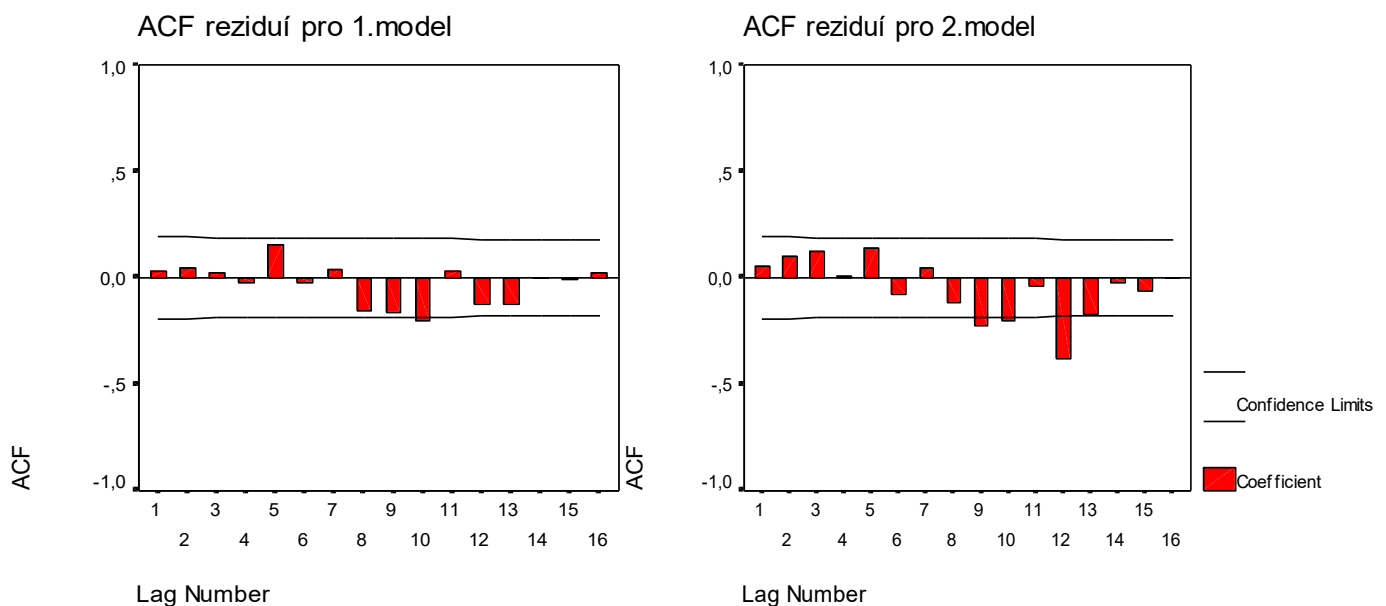
Protože se hodnota blíží k nule, a nepohybuje se tedy okolo hodnoty 2, zamítáme hypotézu o nezávislosti náhodných poruch.

C) Výběrová autokorelační funkce ACF se rovněž používá pro posouzení, zda rezidua jsou

autokorelována. Funkce je definována:
$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad k = 0, 1, \dots$$

Pokud hodnoty výběrové ACF nepřekračují meze 95% intervalů spolehlivosti, je možné předpokládat, že rezidua mají charakter bílého šumu.

Příklad 4. Následující grafy zachycují autokorelační funkce, pro dva různé modely. Na základě těchto grafů určíme, který model je pro danou časovou řadu vhodnější.



Řešení: Hodnoty ACF u 1.modelu nepřekračují meze 95% intervalů spolehlivosti a proto je možné předpokládat, že rezidua v 1.modelu mají charakter bílého šumu. Tento model bude zřejmě pro analýzu časové řady vhodnější.

Případy D,E,F lze použít pouze pro trendové funkce, které jsou lineární v parametrech a jejichž parametry lze odhadnout metodou nejmenších čtverců.

D) Index determinace vychází z rozkladu
$$\sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 + \sum_{t=1}^n (\hat{y}_t - \bar{y})^2,$$

tj. rozkladu celkového součtu čtverců na reziduální součet čtverců a regresní součet čtverců.

Celková variabilita časové řady se skládá z variability nevysvětlené (variability reziduí) a variability vysvětlené zvoleným modelem. *Index determinace* je definován jako podíl

$$\text{variability vysvětlené k variabilitě celkové tj. } R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}.$$

Odmocnina z indexu determinace se nazývá *index korelace*. Index determinace (index korelace) nabývá hodnot od nuly do jedné. Jestliže je roven nule, zvolený model nevysvětluje žádnou variabilitu časové řady, je-li roven jedné, model vysvětluje celou variabilitu zkoumané časové řady. Tyto případy prakticky nenastanou, ale lze říci, že trendová funkce je tím vhodnější, čím je její index determinace (index korelace) bližší jedné. Nedostatkem této míry je její závislost na počtu parametrů zvolené funkce.

Tento nedostatek odstraňuje *adjustovaný (modifikovaný) index determinace*:

$$\text{adj}R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p}, \text{ kde } p = k + 1 \text{ je počet parametrů dané trendové funkce,}$$

k je počet regresorů.

E) Pomocí *t-testů* se testuje hypotéza, že jistý parametr zvolené trendové funkce je roven nule proti alternativní hypotéze, že je různý od nuly. Testové kritérium má v čitateli odhad daného parametru a ve jmenovateli odhad směrodatné chyby tohoto odhadu parametru.

$$\text{Testové kritérium } t = \frac{\hat{\beta}_i}{s\beta_i} \approx t(n - p), \quad i = 0, 1, \dots, k$$

Vysoké absolutní hodnoty testových kritérií svědčí ve prospěch alternativních hypotéz, že parametry jsou různé od nuly.

F) Dalším kritériem je *F-test*, kterým se testuje vhodnost zvolené trendové funkce.

Testové kritérium F je podíl vysvětlené variability a nevysvětlené variability:

$$F_{p-1, n-p} = \frac{\frac{S_T}{p-1}}{\frac{S_R}{n-p}} = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \hat{y}_t)^2} \cdot \frac{n-p}{p-1} \approx F(p-1, n-p),$$

kde p je počet parametrů v trendové funkci, $p - 1$ a $n - p$ je počet stupňů volnosti.

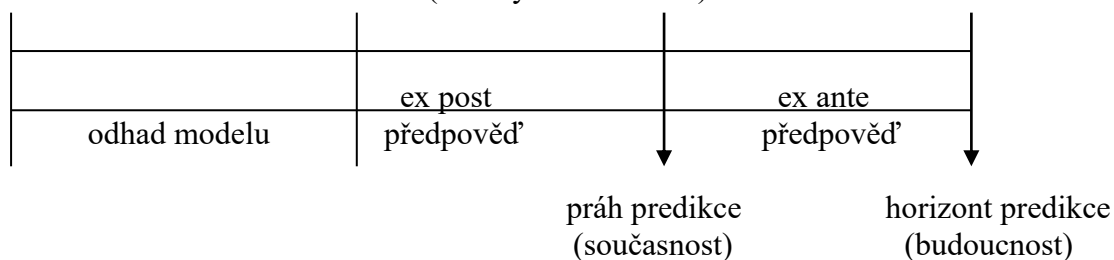
Je-li nulová hypotéza pravdivá, tj. jestliže není daný model vhodný, hodnota testového kritéria by měla být blízka nule. Vysoké hodnoty testového kritéria vedou k zamítnutí nulové hypotézy. Vysoké jsou takové hodnoty testového kritéria, které jsou vyšší než $(1 - \alpha)100\%$ kvantil rozdělení F .

b) Extrapolační kritéria - jsou založena na principu, že se zkoumaná časová řada nejprve zkrátí a na základě této zkrácené řady se provede volba trendové funkce, odhad jejích parametrů a ověření její vhodnosti. Poté se zvolená trendová funkce použije pro výpočet předpovědí, jenž se následně porovnají se skutečnými hodnotami časové řady. Přitom lze použít různé míry charakterizující tuto přesnost „ex post“.

2.4 Metody konstrukce předpovědí časových řad

Mezi nejužívanější metody statistické prognózy patří metody extrapolace časových řad, metody regresní analýzy, strukturální analýzy a metody založené na systémovém přístupu a komplexních modelech. V dalším textu se omezíme jen na metody extrapolace časových řad. Předpovědi se rozumí kvantitativní odhad budoucí hodnoty časové řady, tento odhad je založen na minulém a přítomném průběhu časové řady. Informace o vývoji časové řady je předávána pomocí jejího modelu. Získaná předpověď by proto měla být porovnána s předpovědi získanými jinými metodami (prognózy expertů). Extrapolační prognózy se používají zejména při krátkodobých předpovědích.

Jsou dva druhy předpovědí: *bodové* (odhadují budoucí hodnotu časové řady jedním číslem) a *intervalové* (číselným intervalem).



Předpovědi „ex post“ se provádějí na základě zkrácené časové řady na období, které je známé. Počítají se proto, aby bylo možné hodnotit kvalitu zvoleného prognostického modelu. Práh predikce je poslední známá hodnota časové řady. Horizont předpovědi je délka období, na které se předpovědi počítají.

Hodnocení přesnosti předpovědi

Mějme časovou řadu y_1, y_2, \dots, y_n ukazatele, jehož vývoj chceme předpovídat. Zkonstruovat extrapolační prognózu znamená provést v časovém okamžiku n odhad hodnoty ukazatele v čase i , kde i je zadaný horizont předpovědi.

Označme tento extrapolační odhad $P_n(i)$. Jde o **bodový odhad** odpovídající teoretické hodnoty Y_{n+i} .

Po provedení měření zjistíme, že tento odhad vede k **celkové následné chybě předpovědi**:

$$\Delta = P_n(i) - y_{n+i},$$

kde y_{n+i} je hodnota, kterou sledovaný ukazatel skutečně nabyl v čase $n + i$.

Je-li $\Delta > 0$, jde o **nadceňující předpověď**, je-li $\Delta < 0$, jde o **podceňující předpověď**.

Chybu předpovědi lze rozložit na dvě složky: $\Delta = \Delta^{(m)} + \Delta^{(p)}$,

kde $\Delta^{(m)} = P_n(i) - Y_{n+i}$ je modelová chyba předpovědi, která charakterizuje tu část chyby, která souvisí s volbou modelu,

$\Delta^{(p)} = Y_{n+i} - y_{n+i}$ je chyba vlastního prognostického modelu, která charakterizuje tu část chyby, které se dopustíme v rámci zvoleného modelu.

Intervalové předpovědi

Konstruovat intervalovou předpověď znamená určit interval $P_n(i) - \Delta_n(i) < y_{n+i} < P_n(i) + \Delta_n(i)$ pro dané n $i = 1, 2, \dots$, kde $\Delta_n(i)$ je přípustná chyba předpovědi. Omezíme se na nejčastější případ lineárního trendu. Pro tento případ lze přípustnou chybu psát ve tvaru

$$\Delta_n(i) = t_{1-\frac{\alpha}{2}}(n-2) \cdot s \cdot Q_n(i),$$

kde s je odmocnina z reziduálního rozptylu $s = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{Y}_t)^2}{n-2}}$,

a veličina $Q_n(i) = \sqrt{(1-R^2) \frac{n(n^2-1)+12i}{(n^2-1)(n-2)}}$, kde R^2 je koeficient determinace.

Příklad 5.

Tabulka zachycuje časovou řadu počtu oprav provedených autoopravnou v letech 1995 až 2007. Odhadněme trend lineární trendovou funkcí a provedme bodovou a intervalovou predikci pro roku 2008 až 2010.

Rok	1995	1996	1997	1998	1999	2000	2001
Počet oprav	1901	2085	2124	2431	2858	3164	3150

Rok	2002	2003	2004	2005	2006	2007
Počet oprav	2963	2746	2986	3103	3287	3488

Řešení: Použijeme transformovanou časovou proměnnou $t' = -6, -5, -4, \dots, 4, 5, 6$. Trend odhadneme trendovou funkcí $\hat{T}_t = 2791,23 + 113,67t'$. Po dosazení $t' = 7, 8, 9$; $n = 13$; $i = 1, 2, 3$ dostaneme bodové předpovědi pro následující tři roky:

$$P_{13}(1) = 2791,23 + 113,67 \cdot 7 = 3587,$$

$$P_{14}(2) = 2791,23 + 113,67 \cdot 8 = 3701,$$

$$P_{15}(3) = 2791,23 + 113,67 \cdot 9 = 3814.$$

Za předpokladu, že zůstanou zachovány základní vývojové tendence z let 1995 – 2007, lze očekávat v roce 2008 počet oprav ve výši 3587, v roce 2009 ve výši 3701 a v roce 2010 ve výši 3814.

Abychom mohli konstruovat intervalové předpovědi, vypočteme následující veličiny:

reziduální součet čtverců $S_R = \sum_{t'=-6}^6 (y_{t'} - \hat{T}_{t'})^2 = 684715,$

celkový součet čtverců $S_y = 3\,036\,326,$

index determinace $R^2 = 1 - \frac{S_R}{S_y} = 0,7745,$

odmocnina z reziduálního rozptylu $s = \sqrt{\frac{S_R}{n-2}} = \sqrt{\frac{684\,715}{13-2}} = 249,49,$

veličina $Q_n(i) = \sqrt{(1-R^2) \frac{n(n^2-1)+12i^2}{(n^2-1)(n-2)}} = \sqrt{(1-0,7745) \frac{13(13^2-1)+12}{(13^2-1)(13-2)}} = 0,518.$

Pro $\alpha = 0,05$ je přípustná chyba odhadu

$$\Delta_{13}(1) = t_{0,975}(11) \cdot s \cdot Q_{13}(1) = 2,202 \cdot 249,5 \cdot 0,5148 = 284.$$

Hledaný interval spolehlivosti je $(3587 - 284; 3587 + 284) = (3303; 3871)$.

S 95% spolehlivostí lze očekávat, že v roce 2008 se bude počet oprav pohybovat mezi 3303 až 3871.

Analogicky vypočteme intervalovou předpověď na:

rok 2009 $(3701 - 290; 3701 + 290) = (3414; 3988)$

rok 2010 $(3814 - 290; 3814 + 290) = (3524; 4104)$.

Následné hodnocení přesnosti předpovědí

Nejčastěji používané míry přesnosti jsou:

A) následná chyba předpovědi $D_t(i) = P_t(i) - y_{t+i}, \quad t = 1, 2, \dots, n, \quad i > 0,$

kde $P_t(i)$ jsou hodnoty predikce určené modelem, y_{t+i} jsou skutečné hodnoty analyzovaného ukazatele.

B) Průměrná chyba předpovědi $\bar{D} = \frac{1}{n} \sum_{t=1}^n D_t(i).$

C) Průměrná čtvercová chyba předpovědi $S_D^2 = \frac{1}{n} \sum_{t=1}^n [P_t(i) - y_{t+i}]^2.$

D) Průměrná hodnota chyby předpovědi $S_D = \sqrt{S_D^2}.$ Chyba je považována za uspokojivou, platí-li vztah $-S_D < D_t(i) < S_D.$

E) Relativní míra předpovědi $\partial_t(i) = \frac{D_t(i)}{y_{t+i}}.$

F) Theilův koeficient nesouladu, který je velmi frekventovanou mírou variability relativních chyb předpovědi. Koeficient je dán vztahem:

$$T^2 = \frac{\sum_{t=1}^m (P_t(i) - y_{t+i})^2}{\sum_{t=1}^m y_{t+i}^2},$$

kde $P_t(i)$ jsou hodnoty predikce určené modelem, y_{t+i} jsou skutečné hodnoty analyzovaného ukazatele.

Uvedená charakteristika nebývá nezáporných hodnot. Nulové hranice nabývá pouze v případě bezchybných prognóz. Čím více se koeficient nesouladu odchyluje od nuly, tím více se soustava hodnocených předpovědí liší od ideálních bezchybných předpovědí. Odmocninu z koeficientu nesouladu lze interpretovat jako relativní chybu předpovědi v procentech.

Příklad 5.

Tabulka zachycuje skutečné hodnoty a hodnoty predikované časové řady počtu uchazečů o práci registrovaných na Úřadech práce v České republice do července do prosince roku 2007. Pro konstrukci předpovědí byla zvolena Box – Jenkinsova metodologie. Ověřte pomocí Theilova koeficientu nesouladu, zda je daná funkce vhodná pro výpočet předpovědí.

Řešení:

t	$P_t(1)$	y_{t+1}	$D_t(1)$	$D_t^2(1)$	y^2_{t+1}
7	460559,2773	469728	-9168,7227	84065475,95	2,20644E+11
8	459381,3342	467264	-7882,6658	62136420,11	2,18336E+11
9	455931,177	458272	-2340,823	5479452,317	2,10013E+11
10	443962,588	445174	-1211,412	1467519,034	1,9818E+11
11	440969,8426	442232	-1262,1574	1593041,302	1,95569E+11
12	458760,5193	457369	1391,5193	1936325,962	2,09186E+11
Součet	X	X	X	156678234,7	1,25193E+12

$$\text{Theilův koeficient nesouladu } T^2 = \frac{\sum_{t=1}^m (P_t(i) - y_{t+i})^2}{\sum_{t=1}^m y_{t+i}^2} = 0,000125149 \Rightarrow T = 0,011187 .$$

Můžeme tedy tvrdit, že při konstrukci extrapoláčních prognóz počtu nezaměstnaných jsme se v průměru dopustili chyby 1,18%.

Z vypočtené hodnoty Theilova koeficientu nesouladu vyplývá, že zvolená metoda je z extrapoláčního hlediska vhodná pro konstrukci předpovědí.

Příklad 7.

Tabulka zachycuje extrapoláční předpovědi tempa růstu produkce v procentech $P_t(1)$, pro $t = 1, 2, \dots, 10$ a skutečné hodnoty tempa růstu produkce v následujícím roce y_{t+1} . Vypočítejte všechny absolutní i relativní míry hodnocení přesnosti předpovědi.

t	$P_t(1)$	y_{t+1}
1	2	1
2	3	3
3	4	6
4	5	7
5	0	-2
6	-2	-3
7	3	4
8	6	6
9	1	3
10	-1	-2

Řešení:

Tabulku doplníme o další sloupce, abychom mohli vypočítat absolutní a relativní míry předpovědi.

t	$P_t(1)$	y_{t+1}	$D_t(1)$	$D_t^2(1)$	y^2_{t+1}	$\partial_t(1)$ v %
1	2	1	1	1	1	100
2	3	3	0	0	9	0

3	4	6	-2	4	36	-33,3
4	5	7	-2	4	49	-28,6
5	0	-2	2	4	4	-100
6	-2	-3	1	1	9	-33,3
7	3	4	-1	1	16	-25
8	6	6	0	0	36	0
9	1	3	-2	4	9	-66,7
10	-1	-2	1	1	4	-50
Součet	X	X	-2	20	173	X

Následné chyby předpovědi $D_t(1)$ jsou vypočteny ve 4. sloupci tabulky.

Průměrná chyba předpovědi je $\bar{D} = \frac{1}{n} \sum_{t=1}^n D_t(i) = \frac{-2}{10} = -0,2$.

Průměrná čtvercová chyba předpovědi $S_D^2 = \frac{1}{n} \sum_{t=1}^n [P_t(i) - y_{t+i}]^2 = \frac{20}{10} = 2$.

Průměrná hodnota chyby předpovědi $S_D = \sqrt{S_D^2} = 1,732$.

Nerovnost $-S_D < D_t(i) < S_D$ není splněna pro $t = 3,4,5,9$ a proto příslušné odhady není možné považovat za uspokojivé.

Relativní chyby předpovědi $\partial_t(i) = \frac{D_t(i)}{y_{t+i}}$ jsou vypočteny v posledním sloupci tabulky.

Největší relativní chyby jsou pro $t = 1,5,9,10$.

Theilův koeficient nesouladu $T^2 = \frac{\sum_{t=1}^m (P_t(i) - y_{t+i})^2}{\sum_{t=1}^m y_{t+i}^2} = \frac{20}{173} = 0,1156 \Rightarrow T = 0,34$.

Lze říci, že při konstrukci extrapoláčích prognóz tempa růstu jsme se v průměru dopustili 34% chyby bez ohledu na znaménko.

Shrnutí

Trendová složka představuje nejdůležitější komponentu analyzované časové řady, a proto popis trendu je jedním z nejdůležitějších úkolů analýzy časových řad. Trendová složka totiž poskytuje rozhodující informaci pro prognózování hodnot časové řady do budoucna. K určení trendové složky používáme dva obecné přístupy: *analytický* a *syntetický*.

Analytický přístup stanovení trendu vychází z předem známých typů trendových funkcí vyznačujících se přítomností parametrů, které je třeba stanovit co nejlépe s ohledem na skutečné hodnoty ukazatele časové řady. Výhodou těchto trendových funkcí je to, že je lze snadno použít pro účely prognózování. Nevýhodou je fakt, že typ trendové funkce musíme stanovit předem na základě externích, mnohdy subjektivních předpokladů a informací. Nejužívanější metodou odhadu neznámých parametrů trendové funkce je *metoda nejmenších čtverců*, s níž jsme se setkali již v Kvantitativních metodách B. Zde jsme tuto metodu aplikovali na speciální typ jednoduché regrese pro data ve formě ekonomické časové řady,

tedy případ, kdy nezávisle proměnnou je čas a závisle proměnnou tvoří sledovaný ekonomický ukazatel. Kromě metody nejmenších čtverců pro nelineární trendové funkce uvedeme v této opoře alternativní *metodu vybraných bodů*.

Adaptivní přístup stanovení trendu spočívá ve vyrovnání odchylek daného ukazatele v časové řadě (tzv. vyrovnání) tak, že získané vyrovnané hodnoty vyjadřují trendový faktor obsažený pouze v časové řadě, nikoliv faktor vložený z vnějšku. Nemusíme proto znát předem typ trendové funkce, což je přednost syntetického přístupu oproti přístupu analytickému. Jeho nevýhodou je naopak obtížnější využití pro prognózování hodnot časové řady. Z existujících metod adaptivního přístupu uvedeme ve 4. kapitole.

Řešené příklady

Příklad 8.

V tabulce jsou uvedeny počty prodaných automobilů v autocentru AAA v letech 2000 až 2007.

Rok	2000	2001	2002	2003	2004	2005	2006	2007
Prodané automobily	120	159	167	175	197	172	199	240

- Trend v prodeji automobilů popište lineární trendovou funkcí.
- Jaký počet prodaných automobilů lze očekávat v roce 2008?
- Stanovte koeficient determinace a na jeho základě určete přiléhavost dat k trendové funkci.

Řešení:

- Zavedeme novou časovou proměnnou t' (viz následující tabulka).

Rok	t'	y_t	t'^2	$y_t t'$	\hat{T}	$(y - \hat{T})^2$	$(y - \bar{y})^2$
2000	-7	120	49	-840	133,818	190,937	3436,891
2001	-5	159	25	-795	146,620	153,264	385,141
2002	-3	167	9	-501	159,422	57,426	135,141
2003	-1	175	1	-175	172,224	7,706	13,141
2004	1	197	1	197	185,026	143,377	337,641
2005	3	172	9	516	197,828	667,086	43,891
2006	5	199	25	995	210,630	135,257	415,141
2007	7	240	49	1680	223,432	274,499	3766,891
Součet	0	1429	168	1077		1629,552	8533,875

Vypočítáme odhady b_0 , b_1 parametrů β_0 , β_1 trendové funkce:

$$T_t = \beta_0 + \beta_1 t', t' = -7, -5, -3, \dots$$

Všechny potřebné hodnoty jsou uvedeny v tabulce:

$$b_0 = \frac{\sum y_t}{n} = \frac{1429}{8} = 178,625, \quad b_1 = \frac{\sum t' y_t}{\sum t'^2} = \frac{1077}{168} = 6,410.$$

Odhadnutá trendová funkce má tvar:

$$\hat{T} = 178,625 + 6,41 t', t' = -7, -5, -3, \dots$$

b. Očekávaný prodej v roce 2008 vypočítáme dosazením t' , které odpovídá roku 2008, do rovnice trendu:

$$\hat{T} = 178,625 + 6,401 \cdot 9 = 236,315.$$

Příklad 9.

V tabulce jsou uvedeny údaje o počtu vyrobených myček nádobí v letech 1999 - 2007.

- Trend ve výrobě tohoto výrobku popište exponenciální trendovou funkcí.
- Vypočítejte koeficient determinace a na jeho základě zhodnoťte „přiléhavost“ dat k trendové funkci.

Rok	1999	2000	2001	2002	2003	2004	2005	2006	2007
Myčky nádobí (tis.ks)	8	9	17	20	38	40	70	101	180

Řešení:

- Hledáme odhady b_0, b_1 parametrů trendové funkce

$$T_t = \beta_0 \beta_1^t.$$

Logaritmováním této rovnice přejdeme k vztahu:

$$\ln T_t = \ln \beta_0 + t \ln \beta_1.$$

Zavedením substituce

$$T'_t = \ln T_t, \quad t' = t, \\ \beta'_0 = \ln \beta_0, \quad \beta'_1 = \ln \beta_1$$

se původní rovnice exponenciálního trendu transformuje na rovnici lineárního trendu.

Zavedeme novou časovou proměnnou t'' a vypočítáme koeficienty b'_0, b'_1 podle vztahů.

Všechny potřebné výpočty jsou uvedeny v tabulce.

$$b'_0 = \frac{\sum y'_i}{n} = \frac{31,4886}{9} = 3,4987, \quad b'_1 = \frac{\sum t'' y'_i}{\sum t''^2} = \frac{23,2315}{60} = 0,3872.$$

Potom

$$b_0 = e^{b'_0} = e^{3,4987} = 33,07, \quad b_1 = e^{b'_1} = e^{0,3872} = 1,47.$$

Rok	t''	y	$y' = \ln y$	t''^2	$t'' y'$	\hat{T}	$(y - \hat{T})^2$	$(y - \bar{y})^2$
1999	-4	8	2,0794	16	-8,3178	7,0285	0,8425	2085,7489
2000	-3	9	2,1972	9	-6,5917	10,3519	1,9904	1995,4089
2001	-2	17	2,8332	4	-5,6664	15,2466	2,8771	1344,6889
2002	-1	20	2,9957	1	-2,9957	22,4558	6,2330	1133,6689
2003	0	38	3,6376	0	0	33,0737	24,3049	245,5489
2004	1	40	3,6889	1	3,6889	48,7122	74,1821	186,8689
2005	2	70	4,2485	4	8,4970	71,7452	2,1345	266,6689
2006	3	101	4,6151	9	13,8453	105,6690	16,3831	2240,1289
2007	4	180	5,1930	16	20,7718	155,6333	654,3364	15959,2689
Součet	0	490	31,4886	60	23,2315		783,2839	25458,0001

Hledaná trendová funkce má tvar

$$\hat{T}_{t''} = 33,07 \cdot 1,47^{t''}, t'' = -4, -3, -2, \dots$$

b. Vypočítáme koeficient determinace. Musíme proto znát hodnotu celkového součtu S_y a reziduálního součtu S_R (viz poslední dva sloupce v tabulce).

Pro výpočet reziduálního součtu je třeba znát odhady teoretické hodnoty $\hat{T}_{t''}$, které obdržíme postupným dosazováním za t'' do rovnice trendu.

Např.:

$$t'' = -4: \quad \hat{T} = 33,07 \cdot 1,47^{-4} = 7,08.$$

Všechny hodnoty \hat{T} i součtu S_y, S_R jsou uvedeny v tabulce. Pro koeficient determinace platí:

$$R^2 = 1 - \frac{S_R}{S_y} = 1 - \frac{783,2839}{25458,0001} = 0,969.$$

Hodnota 0,969 říká, že data se těsně přimykají k trendové funkci. Můžeme tvrdit, že tímto modelem je vysvětleno 96,9% celkové variability.

Příklady k procvičení

Ano či ne?...

- 1) Časová řada se dá rozložit na čtyři složky.
- 2) Sezónní složka popisuje nesystematický charakter časové řady.
- 3) Koeficienty lineárního trendu časové řady se odhadují metodou největších čtverců.
- 4) Metoda nejmenších čtverců je založena na minimalizaci chyb.
- 5) Výběrová autokorelační funkce se používá pro posuzování cyklické složky.

Doplňte...

- 6) Mezi trendové funkce patří např.,,,
- 7) Logistický trend má základních vývojově odlišných fází cyklu.
- 8) Mezi charakteristiky, které následně hodnotí přesnost předpovědí patří mimo jiné i koeficient nesouladu.
- 9) Sezónní složka popisuje periodické kolísání v časové řadě během
- 10) Jestliže je časová řada tvořena součtem jednotlivých složek, jedná se o dekompozici časové řady.

Řešte...

11) Počátkem kalendářního roku byl do prodeje zkušebně zaveden nový typ výrobku. O tržbách za jeho prodej (v tis.Kč) máme dispozici tyto údaje:

Pořadí	Měsíc	Tržba
1	leden	26
2	únor	39
3	březen	56
4	duben	83
5	květen	121
6	červen	177
7	červenec	259
8	srpen	380

Předpokládejme, že dosavadní trend vývoje potrvá ještě nejméně šest měsíců (cena výrobku se přitom nebude měnit). Pomocí výpočtu prvních, druhých, třetích diferencí a koeficientu růstu nejděte vhodnou trendovou funkci, odhadněte její parametry, znázorněte původní a vyrovnaná data graficky a odhadněte tržbu za měsíc říjen.

12) Průměrnou měsíční produkci ocelářského výrobního podniku v letech 2003 – 2007 lze popsat časovou řadu 2t; 4t; 3t; 5t; 6t.

a) Znázorněte tuto časovou řadu graficky.

b) Vyrovnajte data lineární trendovou funkcí a znázorněte ji graficky. Použijte transformovanou časovou proměnnou t' . Řešte úlohu bez počítače i s počítačem.

c) Vypočítejte vyrovnané hodnoty pro rok 2003 a 2006 a zakreslete je do grafu.

d) Proveďte předpověď průměrné měsíční produkce pro rok 2012.

13) Následující časová řada obsahuje údaje o produkci textilního podniku v měsících březen až říjen roku 2007 (v Kč):

121 418; 120 401; 124 124; 125 873; 127 002; 129 188; 132 387; 134 200.

Sestavte rovnici trendové přímky a určete bodovou i 90%-ní intervalovou předpověď produkce na měsíc listopad a prosinec 2007.

14) Dopravní firma vykázala v letech 1999 – 2007 své přepravní výkony v tis.tkm:

200; 215; 210; 250; 245; 230; 210; 220; 225.

Vyrovnajte časovou řadu lineárním trendem, odhadněte přepravní výkon firmy v roce 2008 a zhodnoťte výstižnost této trendové funkce koeficientem determinace.

Klíč k řešení

Ano či ne?...

- 1) ano
- 2) ne
- 3) ne
- 4) ano
- 5) ne

Doplňte...

- 6) konstantní trend, lineární trend, kvadratický trend, exponenciální trend, logistický trend
- 7) pět
- 8) Theilův
- 9) jednoho kalendářního roku
- 10) aditivní

Řešte...

11) exponenciální funkce; 815,80 Kč

12) b) $4+0,9t'$; c) 2,2t; 4,9t; d) 10,3t

13) $117\,946,004 + 1972,917t$, listopad 2007 = 135 702,3 Kč; prosinec 2007 = 137 675,2 Kč;
listopad 2007 = <133205,2; 138199,4>; prosinec 2007 = <134999,5; 140350,9>

14) $1,58.t' + 222,7$; rok 2008 = 230,68 tkm; koeficient determinace $R^2 = 0,068$