

Univerzita Karlova v Praze
Matematiko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Kateřina Boková

Predikce časových řad

Katedra teoretické informatiky a matematické logiky

Vedoucí bakalářské práce: Mgr. Martin Pilát, Ph.D

Studijní program: Informatika
Studijní obor: obecná informatika

Praha 2014

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V dne.....

podpis

Název práce: Predikce časových řad

Autor: Kateřina Boková

Katedra : Katedra teoretické informatiky a matematické logiky

Vedoucí bakalářské práce: Mgr. Martin Pilát, Ph.D

Abstrakt: V předložené práci podáváme přehled metod pro modelování a predikci časových řad. Popisujeme jak dekompoziční metody a metody založené na Boxově-Jenkinsově metodologii, tak metody využívající postupů z oblasti výpočetní inteligence, především neuronové sítě. Popis metod je veden především z algoritmického hlediska – jsou uvedeny a odvozeny způsoby, jak nastavovat jejich parametry.

Součástí práce je software, který umožňuje jednotlivé metody aplikovat na časové řady, experimentovat s nimi a porovnávat je.

Klíčová slova: časové řady, predikce, neuronové sítě, regresní analýza

Title: Time series prediction

Author: Kateřina Boková

Department: Department of Theoretical Computer Science and Mathematical Logic

Supervisor: Mgr. Martin Pilát, Ph.D.

Abstract: In this present work, we provide an overview of methods for time series modelling and prediction. We describe methods based on decomposition as well as methods based on the Box-Jenkins methodology. Moreover, we also discuss methods based on the ideas from computational intelligence – mainly neural networks.

The description of the methods is focused on the algorithmic aspects – we derive the ways in which the parameters of the models are set.

The work also contains a software, which allows the user to apply the described methods to given time series and compare them among each other.

Keywords: time series, prediction, neural networks, regression analysis

Obsah

1. Základní pojmy.....	8
1.1. Problémy časových řad.....	9
1.2. Předpovědi časových řad.....	10
1.3. Přístupy k analýze časových řad.....	11
2. Dekompoziční metody.....	12
2.1. Klasická dekompozice.....	12
1. Aditivní dekompozice.....	13
2. Multiplikační dekompozice.....	13
2.2. Řady tvořené trendem a reziduální složkou.....	13
2.2.1. Metoda nejmenších čtverců.....	14
2.2.2. Nejběžnější matematické křivky.....	14
1. Konstantní trend.....	14
2. Lineární trend.....	15
3. Kvadratický trend.....	15
4. Exponenciální trend.....	16
5. Modifikovaný exponenciální trend.....	17
6. Logistický trend.....	17
7. Gompertzův trend.....	18
2.3. Adaptivní metody.....	18
2.3.1. Metoda klouzavých průměrů.....	19
2.3.2. Exponenciální vyrovnání.....	20
1. Jednoduché exponenciální vyrovnání.....	20
2.3.3. Dvojitě exponenciální vyrovnání.....	21
2.4. Sezónní složka.....	23
2.4.1. Odstranění sezónnosti pomocí klouzavých průměrů.....	23
1. Aditivní dekompozice.....	24
2. Multiplikační dekompozice.....	24
2.5. Odhalení cyklické složky.....	24
3. Boxova-Jenkinsova metodologie.....	26
3.1. Stacionarita.....	26
3.2. Modely.....	28
3.2.1. Proces klouzavých součtů MA.....	29
3.2.2. Autoregresní proces AR.....	30
3.2.3. Smíšený proces ARMA.....	30
3.3. Nestacionární řady.....	32
3.3.1. Integrovaný smíšený proces ARIMA.....	32
3.4. Předpovědi.....	33
4. Neuronové sítě.....	34
5. Uživatelská dokumentace.....	37
5.1. Spuštění.....	37
5.2. Načtení řady.....	37
5.3. Editace záložek.....	39
5.4. Editace řad a výběr metody.....	39
6. Závěr.....	44

Úvod

V dnešní dynamické době, kdy každý chce mít náskok před ostatními, by bylo velkou výhodou znát budoucnost a v důsledku toho se umět ve správný okamžik správně rozhodnout. Udělat správné obchodní rozhodnutí nebo dobře investovat peníze a díky tomu zbohatnout. Predikce časových řad, jak už název napovídá, se zabývá právě budoucím chováním časové řady. V podstatě se pokusíme zkonstruovat předpovědi pro časovou řadu. Ale jak to uděláme? A budou tyto předpovědi opravdu spolehlivé? Jedním ze způsobů, jak předpovídat budoucí chování řady je podívat se na časově starší hodnoty predikované řady a snažit se vypožorovat nějaké vlastnosti či vzory chování. Ne nadarmo se říká, že se minulost opakuje. Z toho vyplývá, že se nejdříve budeme snažit nějakou časovou řadu zanalyzovat a pak až předvídat to, jak by se mohla chovat v budoucnosti.

V tomto textu se zpočátku budeme zabývat analýzou a různými vlastnostmi časových řad. Dále se pokusíme pomocí analýzy určit konkrétní metodu vhodnou pro danou časovou řadu tak, aby naše předpovědi byly co nejpřesnější. Poté, co se nám podaří učinit nějaké předpovědi, bychom určitě chtěli znát naši míru úspěšnosti.

Cílem mé bakalářské práce je vytvořit program, který nabídne uživateli načtení a grafické zpracování časové řady. Umožní mu načtenou časovou řadu dále upravovat pomocí různých filtrů a predikovat její budoucí chování pomocí matematických modelů, ale i pomocí metod založených na principu neuronových sítí.

1. Základní pojmy

V této kapitole se seznámíme se základními pojmy, které budeme v následujícím textu používat.

Definice: *Časová řada* je libovolná posloupnost dat y_1, y_2, \dots, y_n chronologicky uspořádaných v čase.

V časových řadách hraje důležitou roli také náhodnost. V našem případě je náhodnost de facto neznalost všech skutečností, jež utvářely hodnotu časové řady v daném momentu jako např. počasí, které ovlivňovalo úrodu. A proto budeme používat také modely, které využívají princip náhodnosti.

Definice: *Náhodný proces* neboli *stochastický proces* je množina $\{Y_t, t \in T\}$ náhodných veličin na stejném pravděpodobnostním prostoru (Ω, A, P) indexovaný pomocí hodnot t z množiny T ($T \subset R$), kde t představuje čas.

Podle parametru t můžeme náhodné procesy dále rozdělit na náhodné procesy ve spojitém čase, kde T je časový interval, nebo na náhodné procesy v diskrétním čase.

Dalším možným rozdělením je dělení podle hodnot náhodných veličin na náhodný proces s diskrétními stavy, na náhodný proces se spojitými stavy a na vícerozměrný náhodný proces, kde Y_t je n -rozměrný náhodný vektor.

Realizací vícerozměrného náhodného procesu jsou pak vícerozměrné časové řady a realizací náhodného procesu s diskrétními či spojitými stavy, kde Y je skalární veličina, je jednoduchá časová řada. Tento text se nadále bude zabývat jen jednoduchými časovými řadami v diskrétním čase, protože s tímto typem časových řad se setkáváme nejčastěji. Vícerozměrné časové řady jsou zobecněním jednoduchých časových řad.

1.1. Problémy časových řad

Jak už bylo výše zmíněno, budeme pracovat s časovými řadami v diskrétním čase, ovšem některé řady jsou ve spojitém čase. Máme dvě možnosti jak dosáhnout diskrétního času. Prvním způsobem je vybrat z každého spojitého intervalu jednu reprezentativní hodnotu, s níž se bude dál pracovat např. poslední naměřenou hodnotu. Druhý způsob je akumulace hodnot, tedy např. sečtení hodnot v daném časovém období.

Jedním z hlavních problémů časových řad je zaznamenávání pomocí času. Celý kalendářní systém je poněkud nepravidelný. Měsíce mají různý počet dnů, každá země má své specifické svátky a v důsledku toho má každá země jiný počet pracovních dnů v měsíci, časová pásma a jiné. Proto byly pro praxi zavedeny jisté standardy, kde měsíc má 30 dnů a rok 360. Tyto standardy jsou používané hlavně v úročení. Nebo můžeme měsíční výdaje přepočítat tak, aby každý měsíc měl 30 dnů tím, že vypočteme průměrnou hodnotu pro jeden den a vynásobíme jí 30.

Definice: *Délka časové řady* je počet pozorování, které vytvářejí danou řadu.

Délka časové řady je také důležitý faktor pro analýzu řad. Např. pro Boxův-Jekinsův přístup potřebujeme minimálně 50 pozorování, aby byl model odpovídající. Je-li délka řady příliš malá, nemusí nám poskytnout dostatek informací pro její analýzu. Naopak při větší délce časové řady se nám může její chování postupně měnit a bude obtížnější danou řadu zanalyzovat a najít pro ní vhodný model. Dalším úskalím příliš dlouhých řad jsou změny hodnot v měření na začátku a konci časové řady způsobené např. inflací.

1.2. Předpovědi časových řad

Naší předpovědi v nějakém budoucím bodě může být konkrétní hodnota nebo rozmezí, kde by se hodnota měla vyskytovat. První typ předpovědi se nazývá bodový a musíme ho brát spíše orientačně. Druhou možností je intervalová předpověď, která nám určí horní a dolní mez v jaké by se předpovídaná hodnota měla nacházet s námi předem určenou pravděpodobností.

Pokud si chceme ověřit předpovědní správnost modelu, tak se často uměle odstraňují některá data, která potom tvoří tzv. *zatajený vzorek*. Předpovědní model se pak zkonstruuje na základě zbylých dat a vypočítají se předpovědi pro časové body zatajeného vzorku. Následně se porovnají výsledky předpovědi se skutečnými výsledky.

Pro kontrolu správnosti zvolené metody můžeme vypočítat chybu e_t naší předpovědi \hat{y}_t , když už známe skutečnou hodnotu y_t . Chybu v čase t spočteme následovně

$$e_t = y_t - \hat{y}_t. \quad (1.2.1)$$

Při výběru správné prediktivní metody a při správném nastavením jejích parametrů by chybová míra měla znázorňovat reziduální složku, tj. složku časové řady, která má vlastnosti šumu (viz. následující kapitola). Pokud chceme naše předpovědi souhrnně porovnat, tak můžeme použít jednu z následujících metod, kde y_t značí hodnotu a \hat{y}_t odhad hodnoty v čase t , n je délka řady a h je počet provedených provedených predikcí.

Součet čtvercových chyb SSE

$$SSE = \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2 = \sum_{t=n+1}^{n+h} e_t^2 \quad (1.2.2)$$

Střední čtvercová chyba MSE

$$MSE = \frac{1}{h} \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2 = \frac{1}{h} \sum_{t=n+1}^{n+h} e_t^2 \quad (1.2.3)$$

Odmocněná střední čtvercová chyba RMSE vytvořená tak, aby vracela chybu ve stejných jednotkách jako předvídaná řada.

$$RMSE = \sqrt{\frac{1}{h} \sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2} = \sqrt{\frac{1}{h} \sum_{t=n+1}^{n+h} e_t^2} \quad (1.2.4)$$

Pro řady s odlehlými pozorováními je vhodná střední absolutní chyba MEA, protože neklade takový důraz na velké chyby.

$$MAE = \frac{1}{h} \sum_{t=n+1}^{n+h} |y_t - \hat{y}_t| = \frac{1}{h} \sum_{t=n+1}^{n+h} |e_t| \quad (1.2.5)$$

Můžeme, také použít metody nezávislé na měřítku řady, jako např. střední absolutní procentní chyby MAPE

$$MAPE = \frac{100}{h} \sum_{t=n+1}^{n+h} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (1.2.6)$$

a jiné.

1.3. Přístupy k analýze časových řad

V následujících kapitolách si popíšeme dva nejběžnější matematické přístupy analýzy časových řad a také budeme časovou řadu modelovat pomocí neuronových sítí. Prvním matematickým přístupem je dekompoziční, který se danou řadu snaží vyjádřit složením jiných jednodušších řad. Druhý přístup je založen na principu Boxovy-Jenkinsovy metodologie, která se snaží vyšetřit závislosti mezi jednotlivými pozorováními.

2. Dekompoziční metody

V této kapitole si popíšeme podle Cipry dva způsoby rozkladů časových řad na jednodušší časové řady, které se dají snadněji předpovědět, a také metody pro predikci řady plynoucí z těchto rozkladů.

2.1. Klasická dekompozice

Asi hlavní vlastností, které nás bude z dlouhodobého hlediska zajímat, je ta, zda hodnoty řady spíše celkově klesají či rostou. Tuto složku budeme nazývat *trend* a budeme ho značit Tr_t . *Sezónní složkou* popíšeme periodické změny v časové řadě během kalendářního roku, které se každý rok opakují, ačkoliv někdy mění svůj charakter. K analýze sezónní složky je zapotřebí mít minimálně pololetní data (tj. data, ve kterých délka intervalu mezi pozorováními je kratší než půl roku). Sezónní složku budeme značit Sz_t . Další složkou, která vykazuje jistou pravidelnost v chování jako trendová nebo sezónní složka je *cyklická složka*. Tato složka zaznamenává růst a pokles v rámci časového úseku s jinou než roční periodou. Její odhalení je obtížné, stejně jako odhalení příčiny jejího vzniku, především kvůli vykazování nepravidelností. Perioda opakování mívá proměnnou délku a obvykle zde bývají i rozdíly mezi sousedními body zvratu.

Po odstranění předešlých tří složek nám zbude směs různých nepravidelných faktorů ovlivňujících řadu. Tuto složku nazýváme reziduální složka a značíme jí E_t . Předpokládáme, že má tvar bílého šumu.

Definice: *Bílý šum* označuje posloupnost nekorelovaných náhodných veličin s nulovou střední hodnotou a konstantním konečným rozptylem.

Nyní už známe všechny potřebné složky časové řady. Jedinou otázkou, která zůstává, je jak se mají poskládat dohromady. Máme dvě možnosti:

1. Aditivní dekompozice

Výsledná řada je součtem všech složek.

$$y_t = Tr_t + Sz_t + C_t + E_t \quad (2.1.1)$$

2. Multiplikativní dekompozice

Výsledná řada vznikne součinem všech složek. trendová složka je měřená v jednotkách řady a ostatní složky jsou bezrozměrné.

$$y_t = Tr_t Sz_t C_t E_t \quad (2.1.2)$$

2.2. Řady tvořené trendem a reziduální složkou

Pokud se nám podařilo časovou řadu očistit od sezónní a cyklické složky (viz další kapitoly), můžeme se nyní pokusit přiřadit křivce tvořené jen trendovou a reziduální složkou

$$y_t = Tr_t + E_t \quad (2.2.1)$$

nějakou známou základní matematickou křivku a dopočítat parametry tak, aby křivka kopírovala tvar predikované časové řady. Předpovědi v tomto případě tvoří hodnoty matematické křivky v čase $t=n+1, t=n+2, \dots$. Přesnost metody se dá zkontrolovat vizuálně či vypočtením chyb. Než budeme rozebírat konkrétní matematické křivky, tak si připomeneme metodu nejmenších čtverců, pomocí níž budeme odhadovat parametry křivek.

2.2.1. Metoda nejmenších čtverců

Metoda nejmenších čtverců je optimalizační metoda, jejímž úkolem je najít odhady b_1, b_2, \dots, b_n parametrů $\beta_1, \beta_2, \dots, \beta_n$, tak aby součet čtverců

$$S = \sum_{t=1}^T (y_t - (\beta_1 + \beta_2 x_{t,2} + \beta_3 x_{t,3} + \dots + \beta_n x_{t,n}))^2 = \sum_{t=1}^T (y_t - x_t \beta)^2 \quad (2.2.2)$$

byl minimální. Nejdříve spočteme parciální derivace podle jednotlivých parametrů.

$$\frac{\rho S}{\rho \beta_1} = 2(y_1 - (\beta_1 + \beta_2 x_{1,2} + \dots + \beta_n x_{1,n})) + \dots + 2(y_n - (\beta_1 + \beta_2 x_{n,2} + \dots + \beta_n x_{n,n}))$$

⋮

$$\frac{\rho S}{\rho \beta_n} = 2x_{1,n}(y_1 - (\beta_1 + \beta_2 x_{1,2} + \dots + \beta_n x_{1,n})) + \dots + 2x_{n,n}(y_n - (\beta_1 + \beta_2 x_{n,2} + \dots + \beta_n x_{n,n}))$$

(2.2.3)

Jelikož hledáme minimum, tak položíme parciální derivace rovny nule a upravíme.

Tím nám vznikne soustava rovnic, jejímž řešením jsou námi hledané odhady.

$$\sum_{t=1}^n y_t = b_1 \sum_{t=1}^n 1 + b_2 \sum_{t=1}^n x_{t,2} + \dots + b_n \sum_{t=1}^n x_{t,n} \quad (2.2.4)$$

⋮

$$\sum_{t=1}^n y_t x_{t,n} = b_1 \sum_{t=1}^n x_{t,n} + b_2 \sum_{t=1}^n x_{t,2} x_{t,n} + \dots + b_n \sum_{t=1}^n x_{t,2} x_{t,n}$$

2.2.2. Nejběžnější matematické křivky

Pomocí metody nejmenších čtverců jsme schopni řadu proložit libovolnou křivkou. Nyní si ukáže křivky, které se používají nejčastěji.

1. Konstantní trend

Tvarem křivky konstantního trendu je přímka rovnoběžná s osou t

$$Tr_t = \beta_0, \text{ kde } t = 1, \dots, n \quad (2.2.5)$$

Konstantní trend použijeme na řadu, kde rozdíly sousedních hodnot $y_{t+1} - y_t$ jsou přibližně nulové. Odhad b_0 parametru β_0 pomocí metody nejmenších čtverců se vyjádří z rovnice

$$nb_0 = \sum_{t=1}^n y_t \quad (2.2.6)$$

Předpověď \hat{y}_t pak nabývá hodnot

$$\hat{y}_t = b_0, \quad \text{pro } t > n \quad (2.2.7)$$

2. Lineární trend

Lineární trend má tvar přímky

$$Tr_t = \beta_0 + \beta_1 t, \quad \text{kde } t = 1, \dots, n \quad (2.2.8)$$

Pokud rozdíly sousedních $y_{t+1} - y_t$ hodnot se pohybují okolo nenulové konstanty, tak bychom měli použít lineární trend. Odhady b_0 a b_1 parametrů β_0 a β_1 pomocí metody nejmenších čtverců dostaneme řešením soustavy rovnic

$$\begin{aligned} b_0 n + b_1 \sum_{t=1}^n t &= \sum_{t=1}^n y_t, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 &= \sum_{t=1}^n t y_t \end{aligned} \quad (2.2.9)$$

Předpověď \hat{y}_t pak vypočteme následovně

$$\hat{y}_t = b_0 + b_1 t, \quad \text{pro } t > n \quad (2.2.10)$$

3. Kvadratický trend

Kvadratický trend reprezentujeme kvadratickým polynomem

$$Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad \text{kde } t = 1, \dots, n \quad (2.2.11)$$

Pohybuje-li se hodnota výrazu $y_{t+2} - 2y_{t+1} + y_t$ okolo nějaké konstanty, tak pro odhad použijeme kvadratický trend. Odhady b_0 , b_1 a b_2 parametrů β_0 , β_1 a β_2 pomocí metody nejmenších čtverců dostaneme jako řešení následující soustavy rovnic

$$\begin{aligned}
b_0 n + b_1 \sum_{t=1}^n t + b_2 \sum_{t=1}^n t^2 &= \sum_{t=1}^n y_t, \\
b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 + b_2 \sum_{t=1}^n t^3 &= \sum_{t=1}^n t y_t, \\
b_0 \sum_{t=1}^n t^2 + b_1 \sum_{t=1}^n t^3 + b_2 \sum_{t=1}^n t^4 &= \sum_{t=1}^n t^2 y_t,
\end{aligned} \tag{2.2.12}$$

Předpověď \hat{y}_t pak nabývá hodnot

$$\hat{y}_t = b_0 + b_1 t + b_2 t^2 \quad \text{pro } t > n. \tag{2.2.13}$$

4. Exponenciální trend

Exponenciální trend se vyjadřuje následovně

$$Tr_t = \alpha \beta^t, \quad \text{kde } t = 1, \dots, n \tag{2.2.14}$$

a je vhodné ho použít na řadu, kde podíl hodnot $\frac{y_{t+1}}{y_t}$ je konstantní.

Logaritmujeme-li rovnici (2.2.14), dostaneme se k rovnici lineárního trendu, kde můžeme vypočítat klasicky oba odhady pro lineární trend. Následným odstraněním logaritmováním dostaneme odhady pro původní rovnici. V praxi se, ale osvědčilo používat pro odhady a a b parametrů α a β místo klasické metody nejmenších čtverců její modifikaci - váženou metodu nejmenších čtverců s vhodně transformovanými hodnotami. Takže místo minimalizace výrazu

$$\sum_{t=1}^T v_t (y_t - \alpha \beta^t), \tag{2.2.15}$$

Minimalizuje zlogaritmovaný výraz

$$\sum_{t=1}^T w_t (\ln y_t - (\ln \alpha + \ln \beta)), \tag{2.2.16}$$

kde w_t jsou váhy závislé na v_t obvykle tímto způsobem

$$w_t = y_t^2 v_t, \quad t = 1, \dots, n, \tag{2.2.17}$$

Výsledkem je odlogaritmované řešení následující soustavy

$$\begin{aligned}
\ln \alpha \sum_{t=1}^n y_t^2 v_t + \ln \beta \sum_{t=1}^n y_t^2 v_t t &= \sum_{t=1}^n (y_t^2 v_t \ln y_t), \\
\ln \alpha \sum_{t=1}^n y_t^2 v_t t + \ln \beta \sum_{t=1}^n y_t^2 v_t t^2 &= \sum_{t=1}^n (y_t^2 v_t t \ln y_t)
\end{aligned} \tag{2.2.18}$$

5. Modifikovaný exponenciální trend

Modifikovaný exponenciální trend, jak je patrné na první pohled, zobecňuje exponenciální trend

$$Tr_t = y + \alpha \beta^t \quad , \text{kde } t = 1, \dots, n \quad (\beta > 0) \quad (2.2.19)$$

Odhady a , b a c parametrů α , β a γ dostaneme jako řešení tří rovnic, které vzniknou rozdělením celé posloupnosti na 3 stejné části (pokud délka posloupnosti není beze zbytku dělitelná třemi, tak vynecháme potřebný počet pozorování na počátku řady). Jednotlivé části následně sečteme. Modifikovaný

exponenciální trend je vhodné použít, když podíl sousedních rozdílů $\frac{y_{t+2} - y_{t+1}}{y_{t+1} - y_t}$

se pohybuje okolo nějaké konstanty.

$$\begin{aligned} \sum_1 y_t &= m\gamma + \frac{\alpha\beta(\beta^m - 1)}{b-1}, & \sum_1 y_t &\text{ je součet první části} \\ \sum_2 y_t &= m\gamma + \frac{\alpha\beta^{m+1}(\beta^m - 1)}{b-1}, & \sum_2 y_t &\text{ je součet druhé části} \\ \sum_3 y_t &= m\gamma + \frac{\alpha\beta^{2m+1}(\beta^m - 1)}{b-1} & \sum_3 y_t &\text{ je součet třetí části} \end{aligned} \quad (2.2.20)$$

Odhady parametrů získáme vyřešením soustavy rovnic (2.2.20). Explicitně vyjádřené odhady mají následující tvar

$$\begin{aligned} b &= \left(\frac{\sum_3 y_t - \sum_2 y_t}{\sum_2 y_t - \sum_1 y_t} \right)^{\frac{1}{m}} \\ a &= \frac{b-1}{b(b^m - 1)^2} (\sum_2 y_t - \sum_1 y_t) \\ c &= \frac{1}{m} \left(\sum_1 y_t - \frac{ab(b^m - 1)}{b-1} \right) \end{aligned} \quad (2.2.21)$$

Předpověď \hat{y}_t pak nabývá hodnot

$$\hat{y}_t = c + ab^t \quad \text{pro } t > n \quad (2.2.22)$$

6. Logistický trend

$$Tr_t = \frac{y}{1 + \alpha \beta^t}, \text{ kde } t=1, \dots, n \quad (\beta > 0, y > 0) \quad (2.2.23)$$

Odhady a , b a c parametrů α , β a γ spočteme pomocí modifikovaného exponenciálního trendu. Ze vzorce je patrné, že logistický trend si lze představit jako převrácenou hodnotu modifikovaného exponenciálního trendu. Postup je pak obdobný jako u modifikovaného exponenciálního trendu, avšak místo

hodnoty y_t se dosadí její převrácená hodnota $\frac{1}{y_t}$. Z toho vyplývá, že jej

použijeme, když je podíl rozdílů převrácených hodnot $\frac{1/y_{t+2} - 1/y_{t+1}}{1/y_{t+1} - 1/y_t}$ konstantní.

Předpověď \hat{y}_t pak nabývá hodnot

$$\hat{y}_t = \frac{1}{c + ab^t}, \text{ pro } t > n \quad (2.2.24)$$

7. Gompertzův trend

Gompertzův trend má následující tvar

$$Tr_t = \exp(\gamma + \alpha \beta^t), \text{ kde } t=1, \dots, n \quad (\beta > 0) \quad (2.2.25)$$

čili

$$\ln Tr_t = \gamma + \alpha \beta^t, \text{ kde } t=1, \dots, n \quad (\beta > 0) \quad (2.2.26)$$

Odhady a , b a c parametrů α , β a γ se provedou obdobně jako u modifikovaného exponenciálního trendu. Místo hodnoty y_t dosadíme do vzorce pro modifikovaný exponenciální trend hodnotu $\exp y_t$. Při přibližně konstantní

hodnotě výrazu $\frac{\ln y_{t+2} - \ln y_{t+1}}{\ln y_{t+1} - \ln y_t}$ volíme Gompertzův trend.

Předpověď \hat{y}_t pak nabývá hodnot

$$\hat{y}_t = \exp(c + ab^t), \text{ pro } t > n \quad (2.2.27)$$

2.3. Adaptivní metody

Aproximace pomocí křivek má jednu nevýhodu. Nedokáže se přizpůsobit změně charakteru řady, ke kterému může dojít. Následující dva druhy metod se řadí do tzv. adaptivních metod, které jsou schopny se přizpůsobit dané změně.

2.3.1. Metoda klouzavých průměrů

Metoda klouzavých průměrů slouží k vyhlazování řady. Lze s ní při správném nastavení parametrů odstranit sezónní nebo cyklickou složku. Základní myšlenka je vytvořit pro každé pozorování vyrovnávací hodnotu pomocí n hodnot před a po tomto pozorování. De facto se snažíme každý časový bod a jeho okolí aproximovat nějakým předem určeným polynomem. Takže vezměme $2n+1$ bodů (je vhodné volit lichou délku posloupnosti kvůli matematickému zjednodušení viz (2.3.4)) časové řady. A jak už bylo výše řečeno, naše první vyrovnávací hodnota bude v bodě $t=n+1$. Vyvažovací hodnota pro bod $t=n+2$ se vypočítá z bodů $y_2 \dots y_{2n+2}$ a tak dále. Aproximace m hodnot, kde $m=2n+1$ v čase t se řádem polynomu r

$$y_{t+\tau}, \quad \text{kde } \tau = -n, \dots, -1, 0, 1, \dots, n \quad (2.3.1)$$

řešíme pomocí metody nejmenších čtverců. Snažíme se najít odhady pro minimalizaci tohoto výrazu

$$\sum_{\tau=-n}^n (y_{t+\tau} - (\beta_0 + \beta_1 \tau + \beta_2 \tau^2 + \dots + \beta_r \tau^r))^2 \quad (2.3.2)$$

Použitím parciálních derivací dostaneme následující soustavu rovnic

$$b_0 \sum_{\tau=-n}^n \tau^j + b_1 \sum_{\tau=-n}^n \tau^{j+1} + b_2 \sum_{\tau=-n}^n \tau^{j+2} + \dots + b_r \sum_{\tau=-n}^n \tau^{j+r} = \sum_{\tau=-n}^n y_{t+\tau} \tau^j, \quad \text{kde } j=0, 1, \dots, r \quad (2.3.3)$$

Nyní tuto poněkud robustní soustavu trochu zjednodušíme. První čeho si všimneme, je to, že pro všechna lichá i platí

$$\sum_{\tau=-n}^n \tau^i = 0 \quad (2.3.4)$$

Protože chceme znát jen prostřední hodnotu, tj. hodnotu v bodě $\tau=0$, k výpočtu hodnoty z polynomu řádu r $b_0 + b_1\tau + b_2\tau^2 + \dots + b_r\tau^r$ nám stačí jen znát hodnotu odhadu b_0 . Tato hodnota je řešením soustavy rovnic vzniklé z v pořadí sudých rovnic, jelikož každá v pořadí lichá rovnice neobsahuje díky (2.3.4) člen s b_0 . Tímto způsobem jsme získali vyrovnávací hodnoty pro skoro celou posloupnost až na prvních a posledních n pozorování. První vyrovnávací hodnoty lze vyjádřit pomocí prvního polynomu v čase s parametry $\tau < 0$ a poslední pomocí posledního polynomu s parametry $\tau > 0$. K těmto výpočtům již musíme dopočítat i ostatní odhady b_1, b_2, \dots, b_r parametrů $\beta_1, \beta_2, \dots, \beta_r$. Tímto způsobem můžeme i predikovat budoucí hodnoty tak, že za parametr τ při aproximaci posledním polynomem dosadíme hodnotu větší než n . Nicméně, čím vzdálenější budoucí předpovědi pomocí této metody vytvoříme, tím méně jsou spolehlivé. Klouzavé průměry se spíše používají k odstranění cyklické či sezónní složky a následně se pak provede aproximace pomocí některé z matematických křivek z části (2.2).

2.3.2. Exponenciální vyrovnání

Při exponenciálním vyrovnávání může pomocí tzv. diskontní konstanty $0 < \beta < 1$ upřednostnit aktuálnější hodnoty před staršími hodnotami.

$$(y_t - \hat{y}_t)^2 + (y_{t-1} - \hat{y}_{t-1})^2 \beta + (y_{t-2} - \hat{y}_{t-2})^2 \beta^2 + (y_{t-3} - \hat{y}_{t-3})^2 \beta^3 + \dots \quad (2.3.5)$$

1. Jednoduché exponenciální vyrovnání

Jednoduché exponenciální vyrovnání se používá pro řady lokálně konstantní.

$$Tr_t = \beta_0 \quad (2.3.6)$$

V tomto případě je odhad $b_0(t)$ závislý na čase t , ze kterého se vypočítává, a získáme ho minimalizací výrazu

$$\sum_{j=0}^{\infty} (y_{t-j} - \beta_0)^2 \beta^j \quad (2.3.7)$$

Po zderivování podle β_0 dostaneme rovnici

$$b_0 = (1 - \beta) \sum_{j=0}^{\infty} y_{t-j} \beta^j .$$

Protože zde mluvíme o řadě lokálně konstantní tak $\hat{y}_t = b_0$.

Nyní ještě převedeme tuto rovnici na rekurzivní vzorec

$$\hat{y}_t = (1 - \beta) y_t + \beta \hat{y}_{t-1} , \quad (2.3.8)$$

který ještě trochu upravíme

$$\hat{y}_t = \alpha y_t + (1 - \alpha) \hat{y}_{t-1} , \quad \text{kde } \alpha = 1 - \beta, 0 < \alpha < 1 \quad (2.3.9)$$

Výhodou této metody je možnost korigovat pomocí parametru α adaptabilitnost metody - čím nižší zvolíme α tím vyhlazenější bude řada. Jako počáteční hodnota parametru \hat{y}_0 se obvykle používá buď hodnota y_1 nebo aritmetický průměr několika prvních hodnot. Počáteční hodnota pro α se v praxi obvykle volí z intervalu $(0, 0.3]$ buď fixně nebo se postupně zkouší všechny tyto hodnoty $0,01 ; 0,02 ; \dots ; 0,3$ a vybere se hodnota s nejmenší chybovou funkcí SSE.

Předpovědi pro $\tau > 0$ je poslední vyrovnávací hodnota pro jednoduché exponenciální vyrovnání v čase t .

$$\hat{y}_{t+\tau}(t) = \hat{y}_t \quad (2.3.10)$$

Pro konstrukci této metody se většinou používá časový bod posledního pozorování.

2.3.3. Dvojitě exponenciální vyrovnání

Dvojitě exponenciální vyrovnání je vhodné pro řady s lokálně lineárním trendem

$$Tr_{t-j} = \beta_0 - j \beta_1 \quad (2.3.11)$$

Pomocí metody nejmenších čtverců minimalizujeme výraz

$$\sum_{j=0}^{\infty} [y_{t-j} - (\beta_0 + \beta_1(-j))]^2 \beta^j \quad (2.3.12)$$

a dostaneme soustavu dvou rovnic o dvou neznámých

$$\begin{aligned} \beta_0 \sum_{j=0}^{\infty} \beta^j - \beta_1 \sum_{j=0}^{\infty} j \beta^j &= \sum_{j=0}^{\infty} y_{t-j} \beta^j, \\ \beta_0 \sum_{j=0}^{\infty} j \beta^j - \beta_1 \sum_{j=0}^{\infty} j^2 \beta^j &= \sum_{j=0}^{\infty} y_t j \beta^j \end{aligned} \quad (2.3.13)$$

Nahrazením sum jejich součtovými vzorci dostaneme následující rovnice

$$\begin{aligned} \beta_0 \frac{1}{1-\beta} - \beta_1 \frac{\beta}{(1-\beta)^2} &= \sum_{j=0}^{\infty} y_{t-j} \beta^j, \\ \beta_0 \frac{\beta}{(1-\beta)^2} - \beta_1 \frac{\beta(1+\beta)}{(1-\beta)^3} &= \sum_{j=0}^{\infty} y_t j \beta^j \end{aligned} \quad (2.3.14)$$

Rovnice následně upravíme tak, abychom měli podobný vzorec jako u jednoduchého exponenciálního trendu.

$$\begin{aligned} \beta_0 - \beta_1 \frac{\beta}{1-\beta} &= (1-\beta) \sum_{j=0}^{\infty} y_{t-j} \beta^j, \\ \beta_0 \beta - \beta_1 \frac{\beta(1+\beta)}{(1-\beta)} &= (1-\beta)^2 \sum_{j=0}^{\infty} y_{t-j} j \beta^j \end{aligned} \quad (2.3.15)$$

Nyní, když celou levou stranu první rovnice nahradíme S_t a celou levou stranu druhé rovnice nahradíme $S_t^{[2]}$, můžeme stejně jako u jednoduchého exponenciálního vyrovnání rovnice přepsat do rekurzivních vzorců

$$\begin{aligned} S_t &= (1-\beta) y_t + \beta S_{t-1} \\ S_t^{[2]} &= (1-\beta) S_t + \beta S_{t-1}^{[2]} \end{aligned} \quad (2.3.16)$$

Po vypočtení S_t a $S_t^{[2]}$ konečně vypočítáme odhady $b_0(t)$ a $b_1(t)$ (značíme $b_0(t)$, protože záleží na čase, ze kterého se odhady počítají) parametrů β_0 a β_1 jako řešení následující soustavy rovnic

$$\begin{aligned} b_0(t) - b_1(t) \frac{\beta}{1-\beta} &= S_t, \\ b_0(t) \beta - b_1(t) \frac{\beta(1+\beta)}{(1-\beta)} &= S_t^{[2]} \end{aligned} \quad (2.3.17)$$

Explicitní vyjádření výsledku je

$$\begin{aligned} b_0(t) &= 2S_t - S_t^{[2]}, \\ b_1(t) &= \frac{1-\beta}{\beta} (S_t - S_t^{[2]}) \end{aligned} \quad (2.3.18)$$

Předpověď $\hat{y}_{t+\tau}$ vytvořená v čase t nabývá hodnot

$$\hat{y}_{t+\tau} = \beta_0(t) + \beta_1(t)\tau, \quad \text{kde } \tau = 1, \dots, n \quad (2.3.19)$$

Počáteční odhady vyjádříme z rovnice (2.3.18) a za $b_0(0)$ a $b_1(0)$ dosadíme odhady parametrů β_0 a β_1 získané z aproximace počátečních hodnot přímkou.

$$\begin{aligned} S_0 &= b_0(0) - \frac{1-\alpha}{\alpha} b_1(0), \\ S_0^{[2]} &= b_0(0) - \frac{2(1-\alpha)}{\alpha} b_1(0), \end{aligned} \quad (2.3.20)$$

2.4. Sezónní složka

Předchozí kapitola se věnovala predikci časových řad tvořených trendovou složkou a šumem. Ale ne vždy máme k dispozici data jen v této formě, tedy data očištěné od sezónnosti a cyklické složky. V této kapitole si ukážeme, jak se dá řada očistit od sezónní i cyklické složky. Sezónní složku nám budou modelovat sezónní faktory I_1, I_2, \dots, I_s , kde s označuje délku sezóny, tj. počet pozorování během jednoho roku.

Nejdříve si ale musíme určit hranici mezi trendovou a sezónní složkou, jelikož není vždy jednoznačná. Pro tento účel si zavedeme *normalizační pravidlo*. Normalizační pravidlo pro aditivní dekompozici vyžaduje součet sezónních faktorů za každý rok rovný nule. Pro multiplikativní dekompozici se součin sezónních faktorů za každý rok musí rovnat jedné.

2.4.1. Odstranění sezónnosti pomocí klouzavých průměrů

Jak už bylo dříve zmíněno, klouzavé průměry mohou zbavit řadu sezónnosti. Popsaná metoda tedy bude založena na principu centrovaných klouzavých průměrů. Centrované klouzavé průměry jsou modifikací klouzavých průměrů sudé délky m , kdy výsledný odhad spadá mezi dvě pozorování. Situaci řešíme tak, že spočteme aritmetický průměr dvou nejbližších klouzavých průměrů sudé délky (dvou meziobdobí).

1. Aditivní dekompozice

Vytvoříme centrované klouzavé průměry \bar{y}_t , které použijeme jako odhad trendu. Odstraníme trendovou složku z řady

$$y_t^* = y_t - \bar{y}_t \quad (2.4.1)$$

a odhadneme sezónní faktory. I_j^* se odhadne jako aritmetický průměr všech j -tých hodnot v každém roce. Dále námi získané odhady centrujeme odečtením aritmetického průměru sezónních faktorů

$$I_j = I_j^* - \bar{I}^* = I_j^* - \frac{I_1^* + I_2^* + \dots + I_s^*}{s}, \text{ kde } j = 1, 2, \dots, s. \quad (2.4.2)$$

Po odečtení odhadů z původní řady, dostaneme řadu očištěnou od sezónní složky

$$\bar{y}_t = y_t - I_j. \quad (2.4.3)$$

2. Multiplikativní dekompozice

Podobně se provádí sezónní očištění i u multiplikativní dekompozice. Zkonstruujeme centrované klouzavé průměry. Odstraníme trendovou složku z řady

$$y_t^* = \frac{y_t}{\bar{y}_t}. \quad (2.4.4)$$

a odhadneme sezónní faktory aplikací geometrického průměru na všechny j -té hodnoty v každém roce. Abychom splnili normalizační pravidlo, tak řadu vydělíme celkovým geometrickým průměrem

$$I_j = \frac{I_j^*}{\widehat{I}^*} = \frac{I_j^*}{\sqrt[s]{I_1^* + I_2^* + \dots + I_s^*}} \quad , \text{kde } j=1,2, \dots, s \quad . \quad (2.4.5)$$

A vydělením dostaneme řadu očištěnou od sezónní složky .

$$\bar{y}_t = \frac{y_t}{I_j} \quad . \quad (2.4.6)$$

2.5. Odhalení cyklické složky

Cyklická složka se obtížně určuje, zvláště je-li v dané řadě skryto více cyklických složek. Pro odhalení cyklické složky bude zapotřebí periodogram a statistické testy.

Definice: *Periodogram* $I(\omega)$ časové řady y_1, y_2, \dots, y_n je funkcí frekvence ω měřených v radiánech za časovou jednotku. Defínovaný následovně

$$\begin{aligned} I(\omega) &= \frac{1}{4\pi} (a^2(\omega) + b^2(\omega)), \quad 0 \leq \omega \leq \pi, \\ a(\omega) &= \sqrt{\frac{2}{n}} \sum_{t=1}^n y_t \cos(\omega t), \\ b(\omega) &= \sqrt{\frac{2}{n}} \sum_{t=1}^n y_t \sin(\omega t), \end{aligned} \quad (2.5.1)$$

Fisherův statistický test periodicity testuje nulovou hypotézu, že daná řada neobsahuje periodickou složku. Takže nulová hypotéza je tvaru

$$H_0: y_t = \mu + \epsilon_t, \quad t=1,2, \dots, n \quad . \quad (2.5.2)$$

Proti nulové hypotéze stojí myšlenka, že řada je tvořena směsí periodických složek

$$y_t = \mu + \sum_{l=1}^k (\alpha \cos(\omega_l t) + \beta \sin(\omega_l t)) + \epsilon_t \quad t=1,2, \dots, n \quad . \quad (2.5.3)$$

Test vychází z hodnot periodogramu v bodech ω_j^*

$$\omega_j^* = \frac{2\pi j}{n}, \quad j=1,2, \dots, m \quad , \text{kde } m = \left\lceil \frac{n-1}{2} \right\rceil \quad (2.5.4)$$

A hodnoty periodogramu spočítáme tímto vzorcem

$$W = \max_{1,2, \dots, m} Y_j = \max_{1,2, \dots, m} \frac{I(\omega_j^*)}{I(\omega_0^*) + I(\omega_1^*) + \dots + I(\omega_m^*)} \quad (2.5.5)$$

Pro rozdělení statistiky W použijeme aproximační vzorec

$$P(W > x) = m(1-x)^{m-1} \quad 0 < x < 1, m < 50 \quad (2.5.6)$$

Splňuje-li hodnota W aproximační vzorec, tak hledanou frekvencí je bod ω_j^* . Po odstranění hodnoty $I(\omega_j^*)$ z periodogramu opakujeme postup. Při neplatnosti aproximačního vzorce skončíme. Necht' máme nalezené všechny odhady frekvencí periodických složek $\omega_1, \dots, \omega_k$, pak z rovnice (2.5.3) vypočteme pomocí metody nejmenších čtverců následující odhady

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n y_t, \quad \hat{\alpha}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(\hat{\omega}_j t), \quad \hat{\beta}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(\hat{\omega}_j t), \quad j=1, \dots, k$$

(2.5.7)

3. Boxova-Jenkinsova metodologie

Boxova-Jenkinsova metodologie přistupuje k analýze časové řady z úplně jiného hlediska, než tomu bylo u dekompozičních metod. Základem této metodologie je prozkoumat vztah mezi jednotlivými pozorováními. Boxova-Jenkinsova metodologie by měla zvládnout predikovat i řady, které jsou dekompoziční metodou obtížně zvládnutelné. Tato metodologie nese v názvu autory monografie Box Jenkins(1970). Hlavním přínosem této monografie bylo vytvoření algoritmické podoby popsanych metod.

Cipra uvádí, že mezi hlavní výhody této metodologie patří to ,že stochastické modely typu ARMA jsou značně flexibilní a tedy použitelné i pro časové řady velmi obecných průběhů a zatím neexistuje „lepší“ rutinní nástroj pro analýzu časově závislých pozorování. Oproti tomu, ale Boxova-Jenkinsova metodologie vyžaduje delší časové řady (jako minimální délka se doporučuje padesát pozorování, to ovšem nebývá problém pro finanční časové řady) a Boxova-Jenkinsova metodologie je v podstatě nerealizovatelná bez počítače vybaveného příslušným softwarem a bez určité instruktáže

3.1. Stacionarita

Jak jsme si řekli v první kapitole, na časovou řadu se můžeme dívat jako na náhodný proces $\{Y_t, t \in T\}$. Ze začátku předpokládejme, že naše řada splňuje podmínku slabé stacionarity. V této kapitole budeme pod pojmem stacionarita vždy rozumět pojem slabá stacionarita.

Definice

Slabě stacionární náhodný proces je invariantní vůči posunům v čase v rámci druhých momentů, tj. pro náhodnou veličinu Y_t platí

$$E(Y_t) = \mu = konst \quad (3.1.1)$$

a pro každé s a t platí

$$\text{cov}(Y_t, Y_s) = E(Y_t - \mu)(Y_s - \mu) = \text{cov}(Y_{t-h}, Y_{s-h}), \text{ pro libovolné } h. \quad (3.1.2)$$

Sousední pozorování v časové řadě mezi sebou většinou vzájemně souvisí, tj. má-li pozorování v čase t hodnotu y_t , tak je dost pravděpodobné, že časově nejbližší hodnoty $\dots, y_{t-1}, y_{t+2}, y_{t+1}, \dots$ nabývají podobných hodnot. Proto si zadefinujeme funkce vyjadřující tyto vztahy pro slabou stacionaritu (střední hodnota μ a rozptyl σ_Y^2 náhodné veličiny Y_t jsou konstantní).

Autokovariační funkce pro zpoždění k je

$$\gamma_k = \text{cov}(Y_t, Y_{t-k}) = E(Y_t - \mu)(Y_{t-k} - \mu) \quad (3.1.3)$$

Autokorelační funkce pro zpoždění k má tvar

$$\rho_k = \frac{\gamma_k}{\sigma_y^2} = \frac{\gamma_k}{\gamma_0}, \quad k = \dots, -1, 0, 1, \dots \quad (3.1.4)$$

Vypočtené korelace můžeme znázornit grafem, tzv. *korelogramem*. Protože autokovariační a autokorelační funkce jsou sudé, tak nám stačí znát jen hodnoty pro $k \geq 0$.

V praxi se autokovariační funkce odhaduje následovně

$$c_k = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad k = 0, 1, \dots, n-1, \quad (3.1.5)$$

kde $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$ je odhad střední hodnoty.

A autokorelační funkce má tvar

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1. \quad (3.1.6)$$

Aby dané odhady byly věrohodné, tak se doporučuje tyto odhady provádět na řadě délky minimálně $n = 50$ a volit $k = \frac{n}{4}$.

Další funkcí, která nám pomůže blíže analyzovat vztah mezi pozorováními je parciální autokorelační funkce. Jejíž odhad je vyjádřen rekurentním vzorcem

$$r_{11}=r_1, \quad r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}, \text{ pro } k > 1, \quad (3.1.7)$$

kde

$$r_{k,j} = r_{k-1,j} - r_{kk} r_{k-1,k-j}, \text{ pro } j = 1, 2, \dots, k-1. \quad (3.1.8)$$

Průběh odhadnutých autokorelačních a parciálních autokorelačních funkcí nám pomáhá identifikovat model (viz následující kapitola). Zejména nás zajímá tzv. bod useknutí k_0 , kdy se korelační hodnota pohybuje blízko nule, pokud takový bod existuje.

3.2. Modely

Základem BoxovyJenkinsovy metodologie je *lineární proces* tvaru

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots, \quad (3.2.1)$$

kde je $\{\varepsilon_t\}$ bílý šum a ψ_1, ψ_2, \dots jsou parametry, které musejí splňovat podmínku

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty, \quad (3.2.2)$$

aby řada (3.2.1) konvergovala a proces byl stacionární s nulovou střední hodnotou.

Zavedeme-li funkci

$$\Phi(z) = \sum_{i=1}^{\infty} \psi_i z^i, \quad (3.2.3)$$

můžeme také říci, že proces je stacionární a řada (3.2.1) konverguje, když $\Phi(z)$ konverguje pro $|z| \leq 1$ v komplexní rovině.

Lineární proces můžeme také vyjádřit i pomocí ostatních náhodných veličin jako

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \dots + \varepsilon_t \quad (3.2.4)$$

za podmínky, že je daný proces invertibilní, tj. $\Pi(z)$ konverguje pro $|z| \leq 1$.

Pro zjednodušení zápisu si zavedeme operátor časového posunu B tímto způsobem

$$B^j Y_t = Y_{t-j} \quad (3.2.5)$$

3.2.1. Proces klouzavých součtů MA

Proces klouzavých součtů $MA(q)$ je v podstatě lineární proces ukončený v bodě q

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \Theta(B) \varepsilon_t \quad (3.2.6)$$

$MA(q)$ je stacionární proces s nulovou střední hodnotou. Jeho rozptyl lze vyjádřit jako

$$\sigma_Y^2 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2 \quad (3.2.7)$$

a autokorelační funkce má tvar

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{(1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2} & \text{pro } k = 1, \dots, q \\ 0 & \text{pro } k > q \end{cases} \quad (3.2.8)$$

$MA(q)$ je invertibilní, tj. $MA(q)$ lze nahradit $AR(\infty)$, když pro kořeny z_1, \dots, z_q $\Theta(z)$ platí, že leží vně jednotkového kruhu v komplexní rovině.

Odhad modelu

Model identifikujeme z odhadnutého korelogramu a parciálního korelogramu. $MA(q)$ má u autokorelační funkce bod useknutí $k_0 = q$ a parciální autokorelační funkce je omezená lineární kombinací geometricky klesajících posloupností a sinusoid různých frekvencí s geometricky klesající amplitudou. Pro lepší odhad bodu useknutí z odhadnutého korelogramu, můžeme námi vybraný bod otestovat pomocí Bartlettovy aproximace

$$|r_k| \geq 2 \sqrt{\frac{1}{n} \left(1 + \sum_{j=1}^{k_0} r_j^2\right)} \quad \text{pro } k > k_0 \quad (3.2.9)$$

Odhady parametrů $\theta_1, \dots, \theta_q$ spočteme pomocí rekurentních vzorců z inovačního algoritmu [1] s proměnnými $\hat{\theta}_{q,1}, \dots, \hat{\theta}_{q,q}$, kde $\hat{\theta}_1 = \hat{\theta}_{q,1}, \dots, \hat{\theta}_q = \hat{\theta}_{q,q}$,

$$\hat{v}_0 = c_0$$

$$\hat{\theta}_{q,q-k} = \hat{v}_k^{-1} \left(c_{q-k} - \sum_{i=0}^{k-1} \hat{\theta}_{q,k-i} \hat{\theta}_{q,q-i} \hat{v}_i \right) \quad , \text{kde } k=0, \dots, q-1 \quad (3.2.10)$$

$$\hat{v}_q = c_0 + \sum_{i=0}^{q-1} \hat{\theta}_{q,q-i}^2 \hat{v}_i$$

a

$$\hat{v}_q = \sigma^2 \quad (3.2.11)$$

3.2.2. Autoregresní proces AR

Autoregresní proces AR(p) má tvar invertovaného lineárního procesu

$$\varepsilon_t = Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = \phi(B) Y_t \quad (3.2.12)$$

AR(p) je vždy invertibilní proces, tj. MA(q) lze vždy nahradit AR(∞), a je stacionární, když pro kořeny z_1, \dots, z_q $\Phi(z)$ platí, že leží vně jednotkového kruhu v komplexní rovině. Za podmínky stacionarity má nulovou střední hodnotou a jeho rozptyl zapíšeme následovně

$$\sigma_Y^2 = \frac{\sigma^2}{(1 - \phi_1 \rho_1 - \dots - \phi_q^2 \rho_p)} \quad (3.2.13)$$

Kovarianční funkce vyjádříme tzv. YuleWalkerovými rovnicemi

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad \text{pro } k=1, \dots, p \quad (3.2.14)$$

Odhad modelu

Autokorelační funkce připomíná lineární kombinaci geometricky klesajících posloupností a sinusoid různých frekvencí s geometricky klesající amplitudou a parciální autokorelační funkce má pod useknutí $k_0 = p$, jenž můžeme otestovat následovně

$$|r_{kk}| \geq 2 \sqrt{\frac{1}{n}} \quad \text{pro } k > k_0 \quad (3.2.15)$$

Odhady parametrů ϕ_1, \dots, ϕ_p můžeme získat z YuleWalkerových rovnic dosazením odhadů r_k za ρ_k a vyřešením soustavy rovnic. Odhad ρ rozptylu bílého šumu vypočteme jako řešení rovnice(3.2.13)

3.2.3. Smíšený proces ARMA

Smíšený proces ARMA(p, q) vznikl spojením předchozích dvou procesů

$$Y_t = \rho_1 Y_{t-1} + \dots + \rho_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-p} + \dots + \theta_q \epsilon_{t-q} \sim \phi(B)y_t = \theta(B)\epsilon_t \quad (3.2.16)$$

Proces ARMA(p, q) je stacionární, když platí podmínka stacionarity pro AR(

p) a ARMA(p, q) je invertibilní, když platí podmínka invertibility pro MA(q). Pokud je ARMA(p, q) stacionární proces, tak má nulovou střední hodnotu a korelační funkci vyjádříme jako

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad \text{pro } k > q \quad (3.2.17)$$

Model AR(p) je speciálním případem modelu ARMA(p, q), kde $q=0$, a model MA(q) je také speciální případem modelu ARMA(p, q) pro $p=0$.

Odhad procesu

Autokorelační funkce po $q-p$ a parciální autokorelační funkce po $p-q$ hodnotách jsou omezené lineární kombinací geometricky klesajících posloupností a sinusoid různých frekvencí s geometricky klesající amplitudou.

Protože k procesu ARMA(p, q) (3.2.16) existuje jeho nekonečná alternativa

$$Y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots \quad (3.2.18)$$

,kde $\sum_{i=0}^{\infty} |\psi_i| < \infty$, tak pro $|z| \leq 1$ získáme rovnici $\Psi(z) = \frac{\Theta(z)}{\Phi(z)}$

$$(1 - \rho_1 z - \rho_2 z^2 - \dots - \rho_p z^p)(\psi_0 + \psi_1 z + \psi_2 z^2 + \dots) = (1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q)$$

(3.2.19)

Porovnáním koeficientů u mocnin z dostaneme následující rovnice

$$\begin{aligned} z^0: \psi_0 &= 1 \\ z^1: \psi_1 &= \theta_1 + \varphi_1 \\ z^2: \psi_2 &= \theta_2 + \varphi_1 \psi_1 + \varphi_2 \\ &\vdots \end{aligned} \quad (3.2.20)$$

obecně zapsáno jako

$$\begin{aligned} \psi_0 &= 1 \\ \psi_j &= \theta_j + \sum_{i=1}^{\min(j,p)} \varphi_i \psi_{j-i}, \quad \text{kde } j=1, 2, \dots \end{aligned} \quad (3.2.21)$$

kde $\theta_j = 0$ pro $j > q$. Odhady inovačního algoritmu $\hat{\theta}_{q,1}, \dots, \hat{\theta}_{q,q+p}$

použijeme jako odhady $\psi_1, \psi_2, \dots, \psi_{p+q}$. Po dosazení získáme

$$\hat{\theta}_{q,j} = \theta_j + \sum_{i=1}^{\min(j,p)} \varphi_i \hat{\theta}_{q,j-i}, \text{ kde } j=1, 2, \dots, p+q \quad (3.2.22)$$

$$\hat{v}_q = \sigma^2. \quad (3.2.23)$$

Řešením (3.2.22) soustavy rovnic pro $j=q+1, \dots, p+q$, kde $\theta_j=0$ dostaneme odhady $\hat{\varphi}_1, \dots, \hat{\varphi}_p$. Odhady $\hat{\theta}_1, \dots, \hat{\theta}_q$ získáme z (3.2.22)

$$\hat{\theta}_j = \hat{\theta}_{q,j} - \sum_{i=1}^{\min(j,p)} \varphi_i \hat{\theta}_{q,j-i}, \text{ kde } j=1, 2, \dots, q. \quad (3.2.24)$$

3.3. Nestacionární řady

Nyní už víme, jak vytvořit model pro stacionární řadu. Bohužel většinou se setkáváme s řadami nestacionárními. Důvody nesplnění stacionarity můžou být následující.

Deterministický trend

Časová řada obsahuje polynomický trend nebo periodickou složku, které musíme před stacionárním modelováním eliminovat.

Stochastický trend

Časová řada ukrývá stochastický trend, tj. jeden nebo více kořenů lineárního procesu leží na jednotkové kružnici (leží-li nějaké kořeny vně jednotkové kružnice, jedná se o nestacionární proces explozivního typu, který se od určitého t začne podobat řadě 2^t).

3.3.1. Integrovaný smíšený proces ARIMA

ARIMA(p, d, q) proces umí modelovat nestacionární časové řady se stochastickým trendem. Využívá k tomu diferencování. Nestacionární řadu se stochastickým trendem lze převést na stacionární řadu právě pomocí diferencování.

$$\begin{aligned}
\Delta Y_t &= Y_t - Y_{t-1} = (1-B)Y_t \\
\Delta^2 Y_t &= \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = (Y_t - Y_{t-1})(Y_{t-1} - Y_{t-2}) \\
&= Y_t - 2Y_{t-1} + Y_{t-2} = (1-B)^2 Y_t \\
&\vdots \\
\Delta^d Y_t &= (1-B)^d Y_t
\end{aligned}
\tag{3.3.1}$$

ARIMA(p, d, q) je v podstatě proces ARMA(p, q) aplikovaný na řadu získanou d -tou diferencí

$$\Phi(B)(1-B)^d Y_t = \Theta(B)\varepsilon_t \tag{3.3.2}$$

Odhad modelu

Parametr diferencování d odhadneme prohlídkou odhadnutých korelogramů pro prvních několik diferencí řady. Pokud korelační funkce velmi pozvolna lineárně klesá, tak to naznačuje přítomnost kořene na jednotkové kružnici a je potřeba provést další diferencí.

Parametry p a q odhadneme stejným způsobem jako u procesu ARMA(p, q).

3.4. Předpovědi

Pro konstrukci použijeme fakt, že

$$\begin{aligned}
\hat{y}_{t+j}(t) &= y_{t+j} \quad \text{pro } j \leq 0, \\
\hat{\varepsilon}_{t+j} &= \begin{cases} 0 & \text{pro } j > 0 \\ \varepsilon_{t+j} = y_{t+j} - \hat{y}_{t+j}(t+j-1) & \text{pro } j \leq 0 \end{cases}
\end{aligned}
\tag{3.4.1}$$

a odhady počítáme postupně rekurentně $\hat{y}_{q+1}(q), \hat{y}_{q+2}(q+1), \dots$. Z těchto odhadů spočítáme odhady $\hat{y}_{q+2}(q), \hat{y}_{q+3}(q+1), \dots$ a tak dále až do posledního pozorování. Počáteční hodnoty nastavíme na nulu $\varepsilon_1 = \dots = \varepsilon_q = 0$.

4. Neuronové sítě

Neuronové sítě představují nedeterministický výpočetní model inspirovaný chováním neuronů v mozku. Základní jednotkou neuronových sítí je *neuron* definovaný jako

$$y = f\left(\sum_{i=1}^n (w_i, x_i) + \Theta\right), \quad (4.1)$$

kde y je výstup neuronu, x_i jsou vstupní hodnoty neuronu, w_i jsou jejich synaptické váhy, Θ je práh neuronu a f je přenosová funkce.

Jednou z nejčastěji používaných přenosových funkcí je sigmodiální přenosová funkce

$$y_j = f(z_j) = \frac{1}{1 + e^{-\lambda z_j}}, \quad \text{kde } z_j = \sum_{i=1}^n y_i w_{ij} \quad (4.2)$$

y_i jsou vstupní hodnoty neuronu j a w_{ij} jsou váhy spojů mezi i -tým a j -tým neuronem. Tato funkce má v každém bodě první derivaci,

$$f'(z_j) = \lambda y_j (1 - y_j), \quad (4.3)$$

kterou budeme později potřebovat pro algoritmus zpětného šíření.

Orientovaným propojením jednotlivých neuronů, určením vstupních a výstupních neuronů a nastavením prahové funkce: neuron $\rightarrow R$ a váhové funkce: spoj $\rightarrow R$ získáme *neuronovou síť*. *Vrstevnatou neuronovou síť* definujeme jako neuronovou síť s orientovaným acyklickým grafem spojů, kde neurony rozdělíme do $m+2$ disjunktních hladin neuronů. První hladina se skládá z tzv. *vstupních neuronů*, jejichž výstupní hodnota se rovná té vstupní. Následuje m vrstev skrytých neuronů, které počítají funkci dle vztahu (4.1). Poslední vrstvu tvoří výstupní neurony.

Stejně jako člověk i neuronová síť prochází *procesem učení*, kdy jsou neuronové síti předkládána trénovací data, tj. vstupní hodnoty a k nim příslušné výstupní hodnoty, a cílem je optimalizovat váhovou a prahovou funkci, tak aby chybová funkce byla minimální na všech trénovacích datech. Chybovou funkci zadefinujeme jako

$$E = \frac{1}{2} \sum_p \sum_j (y_{j,p} - d_{j,p})^2, \quad (4.4)$$

kde p označuje vzory, j je počet výstupních neuronů, $y_{j,p}$ výstupní hodnota a $d_{j,p}$ je požadovaná výstupní hodnota.

Neuronové síť se tradičně učí algoritmem zpětného šíření. Jeho postup je následující. Napřed zvolíme náhodně počáteční váhy. Vložíme trénovací data a aktualizujeme hodnoty váhové a prahové funkce proti směru gradientu chybové funkce vyjádřené jako funkce vah sítě. Funkce se postupně upravují od výstupní ke vstupní vrstvě neuronů. Označme $w_{ij}(t)$ váhu spoje z i -tého do j -tého neuronu v iteraci učení t , pak hodnotu v iteraci $t+1$ vyjádříme jako

$$w_{ij}(t+1) = w_{ij}(t) + \Delta_E w_{ij}(t) \quad (4.5)$$

kde $\Delta_E w_{ij}(t)$ je přírůstek váhy $w_{ij}(t)$ minimalizující chybovou funkci E .

Pro výstupní vrstvu přírůstek váhy odhadneme tímto způsobem

$$\begin{aligned} \Delta_E w_{ij} &\simeq -\frac{\partial E}{\partial w_{ij}} = -\frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial z_j} \frac{\partial z_j}{\partial w_{ij}} = -\frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial z_j} \frac{\partial \sum_{i'} w_{i'j} y_{i'}}{\partial w_{ij}} = -\frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial z_j} y_i = \\ &= -\frac{\partial E}{\partial y_j} f'(z_j) y_i = -(y_j - d_j) f'(z_j) y_i = \delta_j y_i \end{aligned} \quad (4.6)$$

a pro skrytou vrstvu

$$\begin{aligned} \Delta_E w_{ij} &\simeq -\frac{\partial E}{\partial w_{ij}} = -\left(\sum_k \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial y_j}\right) \frac{\partial y_j}{\partial z_j} y_i = -\left(\sum_k \frac{\partial E}{\partial z_k} \frac{\partial \sum_{j'} w_{j'k} y_{j'}}{\partial y_j}\right) \frac{\partial y_j}{\partial z_j} y_i = \\ &= -\left(\sum_k \frac{\partial E}{\partial z_k} w_{jk}\right) \frac{\partial y_j}{\partial z_j} y_i = -\left(\sum_k \delta_k w_{jk}\right) f'(z_j) y_i = \delta_j y_i \end{aligned} \quad (4.7)$$

Pomocí (4.3) souhrnně vzorce zapíšeme v podobě

$$w_{ij}(t+1) = w_{ij}(t) + \Delta_E w_{ij}(t) = w_{ij}(t) \alpha \delta_j y_i + \alpha_m (w_{ij}(t) - w_{ij}(t-1)), \quad (4.8)$$

kde

$$\delta_j = \begin{cases} (d_j - y_j) \lambda y_j (1 - y_j) & \text{pro výstupní neurony} \\ (\sum_k \delta_k w_{jk}) \lambda y_j (1 - y_j) & \text{pro skryté neurony} \end{cases} \quad (4.9)$$

kde λ určuje strmost přenosové funkce, k je index pro neurony z vrstvy nad neuronem j , z_j potenciál a δ_j chyba neuronu j , α je parametr učení a α_m moment učení.

5. Uživatelská dokumentace

5.1. Spuštění

Po spuštění programu uživatel spatří formulář se záložkami. Na každé záložce je zobrazen seznam řad, jejich grafické znázornění.

5.2. Načtení řady

Časovou řadu načteme kliknutím na položku Soubor v horním menu a následně zvolíme položku Otevřít. Nebo můžeme použít klávesovou zkratku Ctr+O. Nyní se nám objeví dialogové okno, v němž vybereme soubor, který chceme zobrazit. Program zpracovává pouze textové dokumenty, tj. dokumenty končící příponou .txt. Program načte dva sloupce hodnot oddělené mezerou, středníkem nebo tabulátorem. V prvním sloupci musí být jen časové hodnoty a ve druhém sloupci by měly být pouze číselné hodnoty, v jiném případě program soubor nenačte.

Parametry načítané řady

Název řady

Nastavení časové osy

Od Roky Hodiny

Měsíce Minuty

Do Dny Sekundy

Úprava časové řady

Aritmetický průměr Lineární interpolace

OK Zrušit

Po úspěšném načtení souboru se záznamy chronologicky setřídí a nyní můžeme pomocí formuláře Parametry načítané řady načtená data upravit. V kolonce Název řady uvedeme námi vybrané pojmenování, které bude reprezentovat načtená data. Ve skupině nastavení časové osy můžeme omezit načtenou řadu nastavením platných časových údajů do kolonek Od a Do. Program potom vybere jen pozorování mezi těmito časy. Tyto kolonky nemusí být vyplněné. Abychom zajistili pravidelnosti pozorování, tak si nejdříve musíme zvolit časové jednotky pro načítanou řadu – výběrem v pravé části skupiny Nastavení časové osy. Výběr Lineární interpolace doplní chybějící pozorování pomocí lineární interpolace a výběr Aritmetický průměr vypočítá aritmetický průměr ze všech pozorování v námi zvolené časové jednotce. Jsou-li zvolené obě možnosti, tak se nejdříve provede aritmetický průměr a až potom lineární interpolace. Těmito výběry zajistíme pravidelná pozorování. Takto upravená řada se zobrazí v aktuálně zobrazené záložce, odpovídá-li časová jednotka řady časové jednotce grafu na záložce, jinak se vytvoří nová záložka, kde bude řada zobrazena.

5.3. Editace záložek

Kliknutím pravým tlačítkem na název záložky vyvoláme kontextové menu. V menu máme na výběr ze tří položek Vložit, Přejmenovat a Smazat.

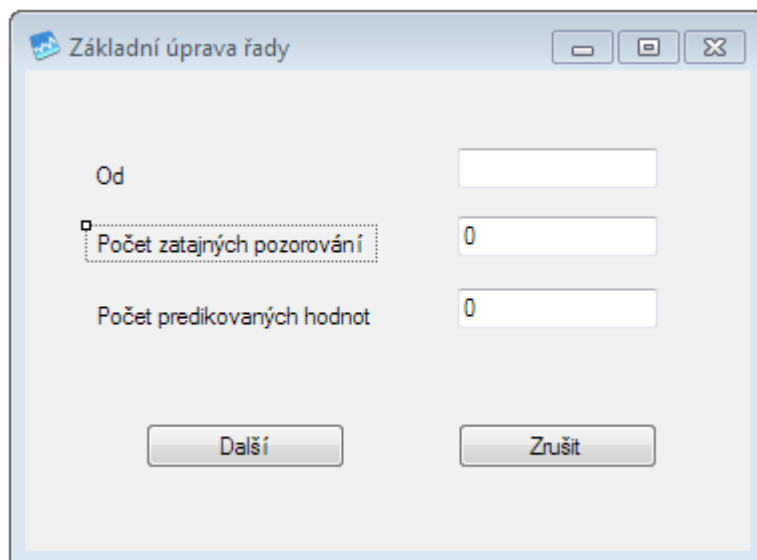
Kliknutím na položku Vložit vytvoříme po zadání názvu novou záložku.

Vybereme-li si Přejmenování, tak vybranou záložku po zadání nového názvu přejmenujeme a výběrem Smazat tuto záložku smažeme.

5.4. Editace řad a výběr metody

Kliknutím pravým tlačítkem na název řady vyvoláme kontextové menu pro řady. Řadu můžeme smazat výběrem položky Smazat nebo po zadání nového názvu přejmenovat výběrem položky Přejmenovat. Kliknutím na položku Uložit otevřeme dialogové okno pro uložení řady.

Výběrem položky Metody vyvoláme formuláře, ve kterých je možné modelovat vybranou řadu.

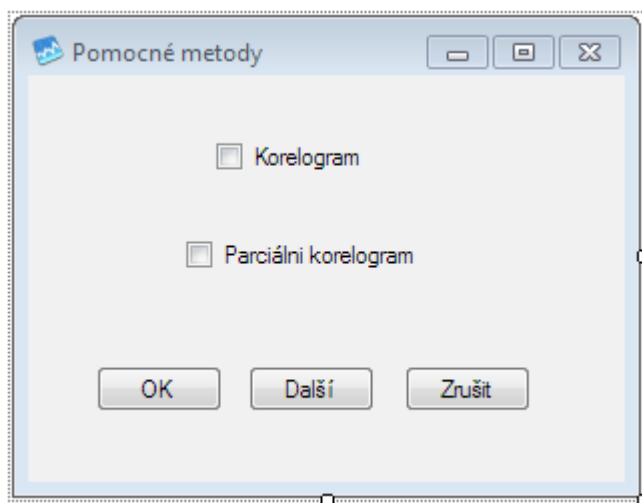


The image shows a dialog box titled "Základní úprava řady" (Basic row editing). It has a standard Windows-style title bar with minimize, maximize, and close buttons. The main area contains three input fields:

- "Od" (From) with an empty text box.
- "Počet zatajných pozorování" (Number of hidden observations) with a text box containing "0".
- "Počet predikovaných hodnot" (Number of predicted values) with a text box containing "0".

At the bottom of the dialog, there are two buttons: "Další" (Next) and "Zrušit" (Cancel).

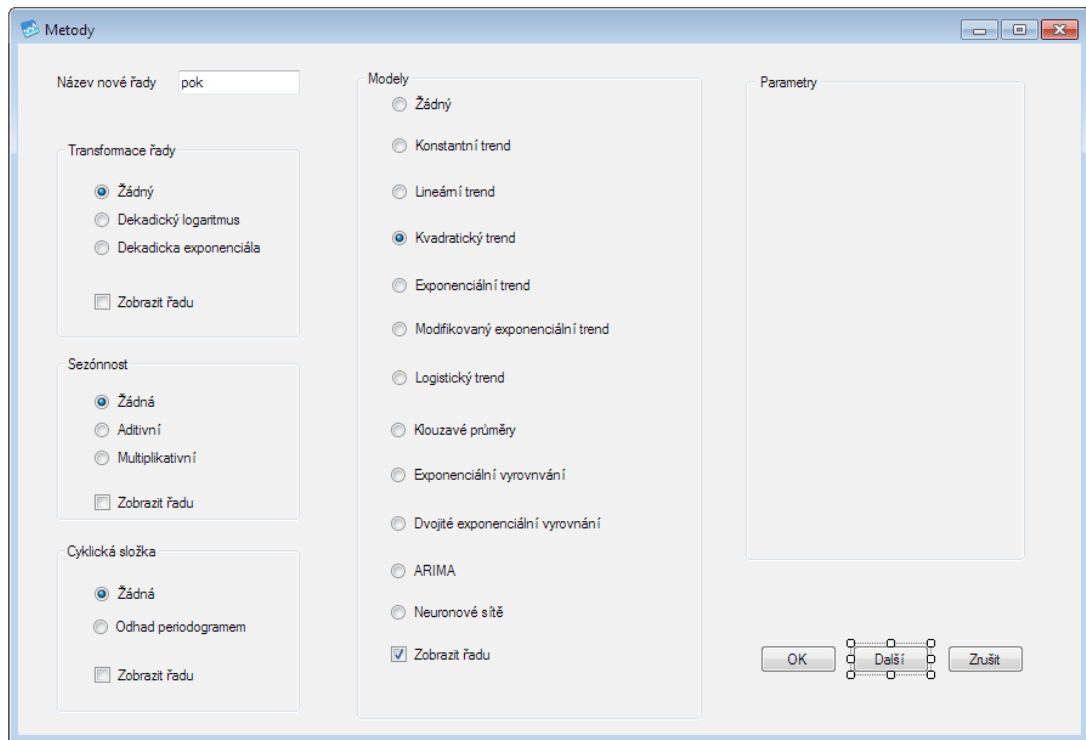
V prvním formuláři nazvaném Základní úprava řady můžeme v položce Od odstranit počáteční hodnoty řady pomocí vloženého data. Není-li nic nastaveno, řada začíná od prvního pozorování. Poté v kolonce Počet zatajených pozorování určíme počet pozorování, které chceme zatajit, abychom následně mohli posoudit předpovědní schopnost modelu a v kolonce Počet predikovaných hodnot nastavíme počet hodnot, které budeme predikovat.



Kliknutím na tlačítko Další zobrazíme druhý formulář s výběrem pomocných metod.

Zaškrtnutím volby Korelogram či Parciální korelogram se v novém okně zobrazí graf hodnot korelační funkce či parciální korelační funkce.

Chceme-li na řadu aplikovat nějaký prediktivní model, tak stiskneme tlačítko Další a vyvoláme tak poslední formulář.



V kolonce Název nové řady zvolíme pojmenování nové řady. Ve skupině Transformace řady zvolíme vhodnou transformaci, tj. všechny hodnoty časové řady se transformuje pomocí dekadického logaritmu nebo dekadické exponenciální funkce. Ve skupině sezónnost z řady odstraníme aditivní (při výběru Aditivní) nebo multiplikativní (při výběru Multiplikativní) sezónnost.

Pokud ve Cyklická složka vybereme Odhad peridogramem, tak se z řady na základě maximálně 50 posledních pozorování odstraní cyklická složka, která vznikla odhadem frekvencí pomocí Fisherova testu. Ve skupině Modely vybereme model, který předpovídá trend. Potřebné další parametry k vybranému modelu se zobrazí ve skupině Parametry.

Nastavení parametrů vyžadují tyto modely:

Klouzavé průměry – tento model potřebuje znát stupeň vyrovnávajícího polynomu a délku aproximované posloupnosti. Zvolíme-li sudou délku posloupnosti, tak se automaticky počítají centrované klouzavé průměry.

Řád se obvykle volí malý a délka posloupnosti by měla odpovídat sezónní nebo cyklické periodě.

Exponenciální vyrovnaní – toto vyrovnaní potřebuje nastavit hodnotu alfa. Hodnota alfa náleží intervalu (0,0.3].

Dvojitě exponenciální vyrovnaní – vyžaduje nastavení hodnoty alfa v intervalu (0,0.3]

ARIMA – vyžaduje nastavení parametru d, p, q. Parametr d určuje stupeň diferencí, parametr p a q řády modelu, odhady konkrétního modelu se spočítají automaticky.

Neuronové sítě – ke správnému fungování neuronových sítí, potřebujeme určit počet vstupních, výstupních neuronů a počet skrytých vrstev. Dále potřebujeme nastavit hodnotu chybovou funkci, při které skončí učení neuronové sítě a maximální počet iterací.

Z vybrané řady se postupně odčítají vymodelované řady a výsledná řada se zobrazí. V případě odstranění všech systematických dekompozičních složek by tato řada měla mít tvar bílého šumu. V opačném případě jsme zvolili špatné metody. Zaškrtnutím ikonky Zobrazit řadu v každé skupině zobrazíme rozšířenou řadu o předpovědní hodnoty vypočtenou vybranou metodou.

6. Závěr

Vytvořená aplikace zobrazuje časové řady a lze na ní pomocí parametrů zvolených v uživatelských formulářích postupně aplikovat některé výše zmíněné metody.

Program používá návrhový vzor model-view-presenter. A jeho knihovni část zvaná model je snadno rozšířitelná o další metody.

- [1] Cipra, T.: Finanční ekonometrie. Ekopress, Praha 2008
- [2] <https://is.muni.cz/el/1431/jaro2006/M0122/M0122.pdf>
- [3] S. Haykin, Neural Networks: A Comprehensive Foundation (2nd Edition). Prentice Hall, July 1998
- [4] <http://ksvi.mff.cuni.cz/~mraz/nn/>