

**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁŘSKÁ PRÁCE**

Martin Grätzer

### **Vybrané přístupy k sezónnímu očišťování ekonomických časových řad**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Radek Hendrych, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční matematika

Praha 2022

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Velice rád bych zde poděkoval svému vedoucímu RNDr. Radkovi Hendrychovi, Ph.D., za odborné vedení mé bakalářské práce, cenné rady a připomínky.

Název práce: Vybrané přístupy k sezónnímu očišťování ekonomických časových řad

Autor: Martin Grätzer

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Radek Hendrych, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce se zabývá problematikou sezónního očišťování ekonomických časových řad a jejich následné predikce. V teoretické části si definujeme časovou řadu a její vlastnosti a popíšeme jednotlivé metody, které budeme využívat: jednoduché přístupy, modelování pomocí kvalitativní proměnné, Holtova-Wintersova metoda, Schlichtova metoda a ETS metody. V praktické části aplikujeme představené metody na reálné ekonomické časové řady. Rozebereme jednotlivé výhody a nevýhody použití dané metody a také se podíváme na chování chybové složky. Následně provedeme jejich komparaci dle schopnosti predikovat následný vývoj dané časové řady.

Klíčová slova: časová řada, sezónnost, predikce, sezónní očišťování

Title: Selected approaches to seasonal adjustment of economic time series

Author: Martin Grätzer

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Radek Hendrych, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis deals with the issue of seasonal adjustment of economic time series and their subsequent predictions. In the theoretical part we define the time series and its properties and describe the individual methods we will use: simple approaches, modeling using a qualitative variable, Holt-Winters method, Schlicht method and ETS methods. In the practical part we apply the presented methods to real economic time series. We will discuss the advantages and disadvantages of using the method and also look at the behavior of the error component. Subsequently, we will compare them according to the ability to predict the subsequent development of a given time series.

Keywords: time series, seasonality, forecast, seasonal adjustment

# Obsah

<b>1</b>	<b>Teoretická část</b>	<b>3</b>
1.1	Časová řada a její dekompozice . . . . .	3
1.2	Jednoduché přístupy k sezónnosti . . . . .	5
1.2.1	Aditivní dekompozice . . . . .	5
1.2.2	Multiplikativní dekompozice . . . . .	6
1.2.3	Predikce . . . . .	7
1.3	Sezónnost modelována pomocí kvalitativní proměnné . . . . .	7
1.4	Holtova-Wintersova metoda . . . . .	9
1.4.1	Aditivní Holtova-Wintersova metoda . . . . .	9
1.4.2	Multiplikativní Holtova-Wintersova metoda . . . . .	10
1.5	Schlichtova metoda . . . . .	11
1.6	ETS metody . . . . .	13
<b>2</b>	<b>Praktická část</b>	<b>20</b>
2.1	Data . . . . .	20
2.2	Software . . . . .	21
2.3	Aplikace jednotlivých přístupů na reálná data . . . . .	21
2.3.1	Jednoduché přístupy k sezónnosti . . . . .	21
2.3.2	Sezónnost modelována pomocí kvalitativní proměnné . . . . .	25
2.3.3	Holtova-Wintersova metoda . . . . .	29
2.3.4	Schlichtova metoda . . . . .	33
2.3.5	ETS Metody . . . . .	34
2.4	Komparace jednotlivých přístupů . . . . .	37
	<b>Seznam použité literatury</b>	<b>41</b>
	<b>Seznam obrázků</b>	<b>42</b>
	<b>Seznam tabulek</b>	<b>43</b>

# Úvod

Tato bakalářská práce se věnuje tématu analýzy ekonomických časových řad. Pomocí představených metod budeme moci časové řady rozložit na jednotlivé složky a taktéž následně predikovat její vývoj. I když lze jednotlivé metody využít i na časové řady bez sezónní složky, tak právě ty s výraznou sezónní složkou budou naším hlavním zájmem.

V teoretické části si představíme koncept časové řady a zadefinujeme jednotlivé složky, ze kterých se skládají. Následně začneme s představováním jednotlivých metod. Prvně si představíme jednoduché metody, které vycházejí spíše z logického postupu než z nějakých matematických teorií. Dále přejdeme z logických úvah do sofistikovanějšího modelování sezónnosti pomocí kvalitativní proměnné a modelování trendu danou funkcí. Poté se budeme věnovat Holtově-Wintersově metodě, která chování časové řady popisuje pomocí systému regresních rovnic. Taktéž se budeme na tuto problematiku dívat z hlediska optimalizace ve Schlichtově metodě. A nakonec si představíme velmi komplexní systém ETS modelů.

V praktické části si představíme ekonomické časové řady, které budeme dále využívat, a taktéž software, ve kterém budeme pracovat. Hlavním tématem této části však bude aplikace představených metod na reálné časové řady, rozebereme jednotlivé výhody a nevýhody použití dané metody. Také se podíváme na splnění předpokladů pro chybovou složku, které jsme definovali v teoretické části. A jako poslední provedeme komparaci jednotlivých metod dle schopnosti predikovat následný vývoj.

# 1. Teoretická část

V první části této práce si definujeme časovou řadu, její složky a způsoby její dekompozice. Dále se budeme věnovat jednotlivým metodám, které si řádně představíme.

Základní informace a strukturu budeme čerpat z knihy Finanční ekonometrie (Cipra, 2013).

## 1.1 Časová řada a její dekompozice

**Definice 1.** Časová řada je posloupnost na sobě závislých chronologicky seřazených náhodných veličin  $\{Y_t\}_{t=1}^T$  definovaných na témže pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ , kde  $T \in \mathbb{Z}$  je délka takové časové řady.

**Definice 2.** Realizaci dané časové řady definujeme jako časovou řadu  $\{Y_t\}_{t=1}^T$  pro jedno fixní  $\omega \in \Omega$ , budeme ji označovat  $\{y_t\}_{t=1}^T$ .

*Poznámka.* V tomto textu budeme pojem časová řada používat i pro její realizaci. Zdali se jedná o časovou řadu nebo její realizaci můžeme poznat z kontextu i ze značení.

Velká část jakýchkoliv dat z ekonomie, přírodních věd či z byznysu je ve formě právě jedné realizace takové časové řady. Jednotlivá pozorování jsou od sebe obvykle časově stejně vzdálená – například dny, týdny, měsíce, čtvrtletí atd.

*Příklad.* Jako příklady časových řad si můžeme uvést,

- Vývoz České republiky v letech 2000 až 2020.
- Vývoj ceny akcie společnosti Apple na newyorské burze v letech 2020 a 2021.
- Teplota za měsíc červen 2021 měřená každou hodinu.

My se zde budeme výhradně zabývat časovými řadami, které mají ekonomický podtext, tedy ekonomické časové řady.

**Definice 3.** Časovou řadu  $\{Y_t\}_{t=1}^T$  můžeme rozložit na složku trendovou  $\{z_t\}_{t=1}^T$ , sezónní  $\{s_t\}_{t=1}^T$  a chybovou  $\{E_t\}_{t=1}^T$  dvojím způsobem. První způsob je aditivní dekompozice:

$$Y_t = z_t + s_t + E_t, \quad \text{pro všechna } t = 1, 2, \dots, T, \quad (1.1)$$

pro kterou platí, že se prvky sezónní a chybové složky uvádí ve stejné jednotce jako je příslušná časová řada. Tou druhou je multiplikativní dekompozice:

$$Y_t = z_t \cdot s_t \cdot (1 + E_t), \quad \text{pro všechna } t = 1, 2, \dots, T, \quad (1.2)$$

kde jsou prvky sezónní a chybové složky bezrozměrné veličiny.

Pro chybovou složku navíc platí, že je to posloupnost nezávislých náhodných veličin  $\{E_t\}_{t=1}^T$  definovaných na témže pravděpodobnostním prostoru  $(\Omega, \mathcal{A}, P)$ . Navíc platí, že jejich střední hodnota je nulová a rozptyl konstantní. Její realizaci označujeme  $\{\epsilon_t\}_{t=1}^T$ .

*Poznámka.* Uvědomme si, že jednotlivé realizace jedné časové řady  $\{Y_t\}_{t=1}^T$  se liší pouze o chybovou složku. Jak trendová, tak i sezónní složka je deterministická.

*Poznámka.* Stejnou dekompozici dle definice 3 lze vytvořit i pro realizaci časové řady  $\{y_t\}_{t=1}^T$ . Pouze chybovou složku  $\{E_t\}_{t=1}^T$  se nahradíme její realizací  $\{\epsilon_t\}_{t=1}^T$ .

Trendová složka vyjadřuje dlouhodobou tendenci vývoje dané časové řady. Je výsledkem faktorů, které dlouhodobě působí stejným směrem, například technologie výroby, demografická situace či podmínky na trhu. Chybová složka vyjadřuje nahodilé a jiné nesystematické výkyvy a u klasických časových řad vyjadřuje i nepřesnosti pozorování, které však nejsou u ekonomických časových řad tak časté.

Avšak v této práci se budeme primárně zabývat složkou sezónní a její eliminací (sezónní očišťování). Sezónní složka je charakteristická svým pravidelným opakováním.

**Definice 4.** *Délka opakování u sezónní složky  $\{s_t\}_{t=1}^T$  časové řady je číslo  $m \in \mathbb{N}$ , pro které platí*

$$s_t = s_{t+m}, \quad \text{pro všechna } t = 1, 2, \dots, T - m.$$

**Definice 5.** *Čísla  $I_1, I_2, \dots, I_m$  nazýváme sezónní faktory sezónní složky  $\{s_t\}_{t=1}^T$  časové řady, platí-li*

$$I_i = s_{i+m \cdot k}, \quad \text{pro všechna } i = 1, \dots, m \text{ a } k \in \mathbb{N}_0.$$

Avšak některé z později představených metod plně nespĺňují definice 4 a 5 a spíše počítají s lokálně adaptivním vývojem sezónních faktorů, který často odpovídá realitě, a tedy  $s_t \approx s_{t+m}$ .

*Příklad.* Jako hypotetický příklad ekonomické časové řady, u které je patrný postupný vývoj sezónních faktorů, můžeme uvést počet prodaných dýní v jednotlivých měsících. Tato řada dříve neměla žádné významné sezónní výkyvy. Nicméně v posledních letech, jak se u nás začíná objevovat americký svátek Halloween, se počet prodaných dýní v měsíci říjen postupně zvyšuje.

Tyto sezónní faktory lze dále využít i při konstrukci předpovědí pro budoucí vývoj.

Někteří autoři přidávají kromě trendové, sezónní a chybové složky ještě složku cyklickou, která je závislá na aktuálním ekonomickém cyklu. Tou se my však zabývat nebudeme a budeme ji uvažovat jako součást trendové složky.

Hlavním rozdílem multiplikativní a aditivní dekompozice je, že v případě aditivní dekompozice velikost sezónní složky na složce trendové nezávisí. Avšak pro multiplikativní dekompozici se sezónní výkyvy zvětšují (respektive zmenšují) pro větší (respektive menší) hodnoty trendové složky.

Dále, abychom vyřešili vzájemnou nejednoznačnost mezi trendovou a sezónní složkou, zavedeme takzvané normalizační pravidlo pro centrování sezónních faktorů.

**Definice 6.** *Mějme aditivní dekompozici časové řady  $\{Y_t\}_{t=1}^T$ , délku opakování sezónní složky  $m$  a sezónní faktory  $I_1, I_2, \dots, I_m$ . Potom definujeme normalizační pravidlo:*

$$I_1 + I_2 + \dots + I_m = 0.$$

**Definice 7.** *Mějme multiplikativní dekompozici časové řady  $\{Y_t\}_{t=1}^T$ , délku opakování sezónní složky  $m$  a sezónní faktory  $I_1, I_2, \dots, I_m$ . Potom definujeme normalizační pravidlo:*

$$I_1 \cdot I_2 \cdot \dots \cdot I_m = 1.$$



## 1.2 Jednoduché přístupy k sezónnosti

Jako první si představíme dvě metody, které jsou svou početní náročností nejjednodušší. Předpokladem pro správné fungování těchto dvou metod je, že se sezónní faktory v jednotlivých sezónách téměř vůbec nemění.

### 1.2.1 Aditivní dekompozice

1. Mějme časovou řadu  $\{y_t\}_{t=1}^T$  a její délku opakování  $m$ .
2. Definujme si  $r = \lfloor m/2 \rfloor$ . Řadu  $\{\bar{y}_t\}_{t=r+1}^{T-r}$  nazveme centrovanými klouzavými průměry.

Pokud je  $m$  liché zkonstruujeme tuto řadu vzorcem:

$$\bar{y}_t^{(m)} = \frac{1}{m} \sum_{i=t-r}^{t+r} y_i, \quad \text{pro všechna } t \in \{r+1, \dots, T-r\}.$$

Pokud je však  $m$  sudé použijeme vzorec:

$$\bar{y}_t^{(m)} = \frac{y_{t-1}^{(m)} + y_t^{(m)}}{2}, \quad \text{pro všechna } t \in \{r+1, \dots, T-r\},$$

kde pro  $y_t^{(m)}$  platí

$$y_t^{(m)} = \frac{1}{m} \sum_{i=t-r+1}^{t+r} y_i, \quad \text{pro všechna } t \in \{r, \dots, T-r\}.$$

3. Nyní však musíme  $\{\bar{y}_t\}_{t=r+1}^{T-r}$  dodefinovat pro  $t$  z intervalu  $\{1, 2, \dots, r-1, r\}$  a  $\{T-(r-1), \dots, T\}$ . Toho lze docílit například zopakováním prvního a posledního spočítaného centrovaného klouzavého průměru, tedy

$$\bar{y}_t^{(m)} = \bar{y}_{r+1}^{(m)}, \quad \text{pro všechna } t \in \{1, 2, \dots, r\},$$

$$\bar{y}_t^{(m)} = \bar{y}_{T-r}^{(m)}, \quad \text{pro všechna } t \in \{T-(r-1), \dots, T\}.$$

Tyto klouzavé průměry lze v rámci této metody považovat za hrubý odhad trendové složky. Díky ní dokážeme řadu očistit od trendu.

4. Definujme řadu  $\{y_t^*\}_{t=1}^T$ , pro kterou platí:

$$y_t^* = y_t - \bar{y}_t^{(m)}, \quad \text{pro všechna } t = 1, 2, \dots, T.$$

Tu můžeme v rámci této metody označovat jako řadu očištěnou od trendové složky. Nyní se tato naše nově zkonstruovaná řada skládá pouze ze složky sezónní a chybové. A můžeme postoupit ke tvorbě odhadů sezónních faktorů.

5. Definujme  $T_j = \left\lceil \frac{T-j+1}{m} \right\rceil$ , potom necentrováné odhady sezónních faktorů vytvoříme vzorcem:

$$I_j^* = \frac{1}{T_j} \sum_{k=0}^{T_j-1} y_{m \cdot k + j}^*, \quad \text{pro všechna } j = 1, \dots, m.$$

A konečně provedeme centrování:

$$I_j = I_j^* - \bar{I}^* = I_j^* - \frac{I_1^* + \dots + I_m^*}{m}, \quad \text{pro všechna } j = 1, \dots, m$$

a dostaneme odhady sezónních faktorů.

6. Řadu  $\{\hat{y}_t^{(m)}\}_{t=1}^T$  nazveme řadou očištěnou od sezónní složky dle metody aditivní dekompozice a vytvoříme ji pomocí následujícího vzorce:

$$\hat{y}_t^{(m)} = y_t - I_j, \quad \text{kde } j = t - m \left\lfloor \frac{t-1}{m} \right\rfloor, \quad \text{pro všechna } t = 1, \dots, T.$$

## 1.2.2 Multiplikativní dekompozice

1-3. První tři kroky stejné jako u aditivní dekompozice.

4. Zkonstruujeme řadu očištěnou od trendové složky  $\{y_t^*\}_{t=1}^T$  vzorcem:

$$y_t^* = \frac{y_t}{\bar{y}_t^{(m)}}, \quad \text{pro všechna } t = 1, 2, \dots, T,$$

5. Definujme  $T_j = \left\lceil \frac{T-j+1}{m} \right\rceil$ , potom necentrováné odhady sezónních faktorů definujeme vzorcem:

$$I_j^* = \frac{1}{T_j} \sum_{k=0}^{T_j-1} y_{m \cdot k + j}^*, \quad \text{pro všechna } j = 1, \dots, m.$$

Z těchto necentrováných odhadů vytvoříme odhady centrované vzorcem:

$$I_j = \frac{I_j^*}{\hat{I}^*} = \frac{I_j^*}{(I_1^* \cdot \dots \cdot I_m^*)^{\frac{1}{m}}}, \quad \text{pro všechna } j = 1, \dots, m.$$

Tím bude splněno normalizační pravidlo.

6. Nyní přistoupíme k očištění původní časové řady od sezónní složky. Řadu  $\{\hat{y}_t^{(m)}\}_{t=1}^T$  definujeme vzorcem:

$$\hat{y}_t^{(m)} = \frac{y_t}{I_j}, \quad \text{kde } j = t - m \left\lfloor \frac{t-1}{m} \right\rfloor, \quad \text{pro všechna } t = 1, \dots, T.$$

Nazveme ji jako řadu očištěnou od sezónní složky metodou multiplikativní dekompozice.

### 1.2.3 Predikce

Predikci časové řady pro následující období je možné vytvořit pomocí predikce dalšího vývoje dané očištěné časové řady definované v předchozích sekcích. Na predikci očištěné časové řady poté aplikujeme sezónní faktory. Tuto predikci vytvoříme proložení této řady funkcí pomocí metody nejmenších čtverců. Pokud se jedná o časovou řadu s lineárním trendem, můžeme použít přímkou. Avšak záleží zcela jistě na povaze dané časové řady.

Pro časovou řadu s velkým množstvím pozorování a několikanásobnou změnou trendu může být velmi složité najít správnou funkci, díky které můžeme data prokládat. V tomto případě lze proložit jednodušší funkcí pouze několik posledních pozorování a zjistit lokální trend, jehož protažením lze predikovat následný vývoj časové řady. Je samozřejmé, že zjištění lokálního chování trendu nám dá často přesnější krátkodobé predikce, což je pravda hlavně pro reálné ekonomické časové řady. Nicméně pokud jsme si jistí, že daná časová řada má nějaké specifické chování, které lze popsat křivou, tak naše dlouhodobé predikce budou výrazně lepší, když budeme vyrovnávat přes všechny či velkou část našich pozorování.

**Definice 8.** Necht  $\{y_t\}_{t=1}^T$  je časová řada s délkou opakování  $m$ . Dále necht  $I_j$  pro  $j = 1, \dots, m$  jsou centrované odhady sezónních faktorů a  $\{\hat{y}_t^{(m)}\}_{t=1}^T$  je časová řada očištěná od sezónní složky. Poté vytvoříme odhady  $\hat{y}_{T+h}^{(m)}$ ,  $h \in \mathbb{N}$  následného chování řady  $\{\hat{y}_t^{(m)}\}_{t=1}^T$  pomocí metody nejmenších čtverců.

Nyní definujeme  $\hat{y}_{T+h}$ ,  $h \in \mathbb{N}$  jakožto predikci vývoje časové řady  $\{y_t\}_{t=1}^T$  pomocí následujícího vzorce:

$$\begin{aligned}\hat{y}_{T+h} &= \hat{y}_{T+h}^{(m)} + I_j, \\ \hat{y}_{T+h} &= \hat{y}_{T+h}^{(m)} \cdot I_j,\end{aligned}$$

kde  $j = T + h - m \lfloor \frac{T+h-1}{m} \rfloor$ . První vzorec použijeme, pokud se jedná o aditivní dekompozici, a druhý, pokud o multiplikativní

## 1.3 Sezónnost modelována pomocí kvalitativní proměnné

Tato metoda je první z regresních metod. Její odlišnost od jednoduché metody aditivní dekompozice spočívá pouze v sofistikovanějším modelování sezónních faktorů pomocí regresních modelů a modelování trendové složky dle námi zvolené funkce.

**Definice 9.** Mějme časovou řadu  $\{Y_t\}_{t=1}^T$  s délkou opakování  $m$  a její aditivní dekompozici  $Y_t = z_t + s_t + E_t$ , pro všechna  $t = 1, 2, \dots, T$ . Trendovou složku  $\{z_t\}_{t=1}^T$  modelujme vhodnou funkcí času  $f(t)$ . Naopak sezónní složku budeme modelovat pomocí kvalitativní proměnné s použitím umělých nula-jedničkových proměnných (tzv. dummy variable), které přiřazují dané proměnné jedničku, nachází-li se časová řada v uvažované sezóně a v ostatních případech nulu, viz tabulka 1.1.

Příslušný regresní model tedy bude v následujícím tvaru,

$$Y_t = f(t) + \alpha_2 x_{t2} + \dots + \alpha_m x_{tm} + E_t, \quad \text{pro všechna } t = 1, 2, \dots, T.$$

$t$	$x_{t2}$	$\cdots$	$x_{ts}$
1	0	$\cdots$	0
2	1	$\ddots$	$\vdots$
$\vdots$	0	$\ddots$	0
$m$	$\vdots$	$\ddots$	1
$m+1$	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabulka 1.1: Umělé proměnné reprezentující sezónnost

**Tvrzení 1.** Mějme časovou řadu  $\{y_t\}_{t=1}^T$  s délkou opakování  $m$ , dále mějme regresní model z definice 9 s tím, že  $f(t) = \beta_0 + \beta_1 t$ , kde  $\beta_0$  a  $\beta_1$  jsou neznámé parametry. Označme potom vektor všech parametrů regresního modelu  $\mathbf{a} = (\beta_0, \beta_1, \alpha_2, \dots, \alpha_m)^T$ , dále označme  $\mathbf{Y} = (y_1, \dots, y_T)^T$ . Následně definujme matici  $\mathbf{F}$ :

$$\mathbf{F} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ 1 & s & \vdots & \ddots & 1 \\ 1 & m+1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & T & \vdots & \vdots & \vdots \end{pmatrix}$$

velikosti  $(T) \times (m+1)$ .

Potom můžeme vektor parametrů  $\mathbf{a}$  odhadnout vektorem  $\hat{\mathbf{a}}$  pomocí metody nejmenších čtverců vzorcem:

$$\hat{\mathbf{a}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{Y}.$$

A získáváme odhady  $\hat{\beta}_0, \hat{\beta}_1, \hat{\alpha}_2$  až  $\hat{\alpha}_m$ .

*Důkaz.* Jednotlivé kroky důkazu jsou popsány v práci Rencher a Schaalje (2000, str.142). □

**Definice 10.** Necht  $\{y_t\}_{t=1}^T$  je časová řada s délkou opakování  $m$ . Mějme regresní model daný dle definice 9 a  $f(t) = \beta_0 + \beta_1 t$ . Dále již mějme odhady parametrů tohoto regresního modelu dle tvrzení 1. Poté definujme

$$\bar{\alpha} = \frac{\hat{\alpha}_2 + \dots + \hat{\alpha}_m}{m}.$$

Nyní vytvoříme odhady sezónních faktorů pomocí následujícího vzorce:

$$\hat{I}_1 = -\bar{\alpha} \quad a \quad \hat{I}_i = \hat{\alpha}_i - \bar{\alpha} \quad \text{pro všechna } i = 2, \dots, m.$$

A zkonstruujeme odhad pro trendovou složku:

$$\hat{z}_t = (\hat{\beta}_0 + \bar{\alpha}) + \hat{\beta}_1 t \quad \text{pro všechna } t = 1, 2, \dots, T.$$

**Definice 11.** Necht  $\{y_t\}_{t=1}^T$  je časová řada s délkou opakování  $m$ . Mějme platný regresní model daný dle definice 9 a  $f(t) = \beta_0 + \beta_1 t$ . A také mějme odhady parametrů tohoto regresního modelu dle tvrzení 1. Poté predikci časové řady o  $h \in \mathbb{N}$  období dopředu můžeme vypočítat jednoduchým dosazením do vzorce:

$$\hat{y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 \cdot (T+h) + \hat{\alpha}_2 x_{(T+h)2} + \dots + \hat{\alpha}_m x_{(T+h)m}.$$

*Poznámka.* Metodu lze modifikovat pro jakoukoliv funkci  $f(t)$ . Pokud bychom se například domnívali, že trend má podobu polynomu třetího řádu, tak by náš regresní model vypadal následovně:  $Y_t = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \alpha_2 x_{t2} + \dots + \alpha_m x_{tm} + E_t$ , pro všechna  $t = 1, 2, \dots, T$ . Tvrzení 1, definice 10 a definice 11 by se změnilly analogicky.

## 1.4 Holtova-Wintersova metoda

Holtova-Wintersova metoda je rozšíření základní Holtovy metody tak, aby byla vedle lokálně lineárního trendu zohledněna také sezónnost.

### 1.4.1 Aditivní Holtova-Wintersova metoda

**Definice 12.** Necht  $\{y_t\}_{t=1}^T$  je časová řada s délkou opakování  $m$ . Potom rekurentní vzorce aditivní Holtovy-Wintersovy metody mají tvar:

$$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}, \\ s_t &= \gamma(y_t - \ell_t) + (1 - \gamma)s_{t-m}, \\ \hat{y}_t &= \ell_t + s_t, \end{aligned}$$

pro všechna  $t = 1, 2, \dots, T$ .

**Definice 13.** Mějme časovou řadu  $\{y_t\}_{t=1}^T$  s délkou opakování  $m$ . Mějme platné rekurentní vzorce z definice 12. Potom odhady pro budoucí vývoj časové řady budou ve tvaru:

$$\hat{y}_{T+h} = \ell_T + b_T \cdot h + s_{T-m+\tau}, \quad \text{pro } h \in \mathbb{N} \text{ a } \tau = (h+1) \bmod m + 1.$$

Realizace rekurentních formulí aditivní Holtovy-Wintersovy metody vyžaduje volbu počátečních hodnot  $\ell_0, b_0, s_{-m+1}, \dots, s_0$  a volbu vyrovnávacích konstant  $\alpha, \beta$  a  $\gamma$ .

Počáteční hodnoty lze určit modelováním sezónnosti pomocí kvalitativní proměnné jako v odstavci 1.3, kde po vyřešení regresního modelu položíme

$$\ell_0 = \hat{\beta}_0, \quad b_0 = \hat{\beta}_1, \quad s_{-m+1} = 0, \quad s_{-m+2} = \hat{\alpha}_2, \quad \dots, \quad s_0 = \hat{\alpha}_m.$$

Pro volbu vyrovnávacích konstant doporučujeme fixní volbu  $\alpha = 0.4, \beta = 0.1$  a  $\gamma = 0.4$ , avšak taktéž lze vytvořit jejich odhad, například pomocí minimalizace sumy druhých mocnin chyb (SSE), což používají jednotlivé softwary. Pokud bychom chtěli konstanty volit sami, tak pro menší  $\alpha$  dostáváme trend, který pomaleji reaguje na jednotlivé výkyvy a změny trendu. Pro velké hodnoty  $\beta$  dostáváme trendovou složku, která celkem rychle mění směr a rychlost změny. A co se týče  $\gamma$ , tak pro nižší hodnoty bude sezónní složka pravidelnější.

## 1.4.2 Multiplikativní Holtova-Wintersova metoda

**Definice 14.** Necht  $\{y_t\}_{t=1}^T$  je časová řada s délkou opakování  $m$ . Potom rekurentní vzorce multiplikativní Holtovy-Wintersovy metody mají tvar:

$$\begin{aligned}\ell_t &= \alpha \left( \frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}, \\ s_t &= \gamma \left( \frac{y_t}{\ell_t} \right) + (1 - \gamma)s_{t-m}, \\ \hat{y}_t &= \ell_t \cdot s_t,\end{aligned}$$

pro všechna  $t = 1, 2, \dots, T$ .

**Definice 15.** Mějme časovou řadu  $\{y_t\}_{t=1}^T$  s délkou opakování  $m$ . Mějme platné rekurentní vzorce z definice 14. Potom odhady pro budoucí vývoj časové řady budou ve tvaru:

$$\hat{y}_{T+h} = (\ell_T + b_T \cdot h) \cdot s_{T-m+\tau}, \quad \text{pro } h \in \mathbb{N} \text{ a } \tau = (h + 1) \pmod{m + 1}.$$

Ve srovnání s aditivní verzí se jednotlivé součty a rozdíly, v nichž figuruje sezónní složka, nahradí součiny a podíly.

Pro realizaci rekurentních formulí multiplikativní Holtovy-Wintersovy metody opět vyžadujeme volbu počátečních hodnot  $\ell_0$ ,  $b_0$ ,  $s_{-m+1}, \dots, s_0$  a volbu vyrovnávacích konstant  $\alpha$ ,  $\beta$  a  $\gamma$ .

**Definice 16.** Necht  $\{y_t\}_{t=1}^T$  je časová řada s délkou opakování  $m$ . Dále si zdefinujeme

$$\begin{aligned}r &= \left\lfloor \frac{T}{m} \right\rfloor, \\ \bar{y}_i &= \frac{1}{m} \sum_{j=1}^m y_{(i-1)m+j}, \quad \text{pro } i = 1, \dots, r.\end{aligned}$$

Počáteční hodnoty pro multiplikativní Holtovu-Wintersovu metodu zvolíme následujícím způsobem:

$$\begin{aligned}b_0 &= \frac{\bar{y}_r - \bar{y}_1}{(r - 1) \cdot m}, \\ \ell_0 &= \bar{y}_1 - \frac{m + 1}{2} \cdot b_0, \\ s_{j-m} &= \frac{1}{r} \sum_{i=0}^{r-1} \frac{y_{j+m \cdot i}}{\bar{y}_{i+1} - \left( \frac{m+1}{2} - j \right) \cdot b_0}, \quad \text{pro } j = 1, \dots, m.\end{aligned}$$

Pro vyrovnávací konstanty znovu doporučíme fixní volbu  $\alpha = 0.4$ ,  $\beta = 0.1$  a  $\gamma = 0.4$ .

## 1.5 Schlichtova metoda

Na rozdíl od předchozích metod Schlichtova metoda přistupuje k problému aditivní dekompozice z hlediska optimalizace. V této podkapitole budeme informace čerpat z článku Ekkeharta Schlichty (Schlicht, 1981).

**Definice 17.** *Mějme časovou řadu  $\{y_t\}_{t=1}^T$  a její aditivní dekompozici  $y_t = z_t + s_t + \epsilon_t$ ,  $\forall t \in \{1, 2, \dots, T\}$ . Definujme*

$$\begin{aligned}\mathbf{y}^\top &= (y_1, y_2, \dots, y_T), \\ \mathbf{z}^\top &= (z_1, z_2, \dots, z_T), \\ \mathbf{s}^\top &= (s_1, s_2, \dots, s_T), \\ \boldsymbol{\epsilon}^\top &= (\epsilon_1, \epsilon_2, \dots, \epsilon_T).\end{aligned}$$

Potom lze aditivní dekompozici přepsat vzorcem:

$$\mathbf{y} = \mathbf{z} + \mathbf{s} + \boldsymbol{\epsilon}.$$

Nyní se podíváme na to, jak by měla intuitivně vypadat naše trendová a sezónní složka.

V první řadě požadujeme, aby trendová složka byla co možná nejvíce hladká křivka. Kvůli tomu si zadefinujeme funkci  $f(\mathbf{z}) : \mathbb{R}^T \rightarrow \mathbb{R}$ , která bude měřit hladkost křivky trendové složky, následujícím způsobem:

$$f(\mathbf{z}) := \alpha \sum_{t=3}^T ((z_t - z_{t-1}) - (z_{t-1} - z_{t-2}))^2, \quad \alpha > 0, \quad (1.3)$$

kde každý sčítanec lze považovat za změnu směru trendu. Potom suma čtverců těchto změn bude přenásobena konstantou  $\alpha$ . Aby náš trend byl co nejvíce hladký, budeme chtít, aby hodnota funkce  $f(\mathbf{z})$  byla co nejmenší.

Od sezónní složky budeme požadovat, aby co nejstabilněji projevovovala sezónní opakování. Navíc také požadujeme, aby součet sezónních faktorů byl téměř nulový. Zde si definujeme funkci  $g(\mathbf{s}) : \mathbb{R}^T \rightarrow \mathbb{R}$ , z jejíž pomocí budeme hodnotit nestabilitu sezónní složky:

$$g(\mathbf{s}) := \beta \sum_{t=m+1}^T (s_t - s_{t-m})^2 + \gamma \sum_{t=m}^T \left( \sum_{\tau=0}^{m-1} s_{t-\tau} \right)^2, \quad \beta \geq 0, \gamma > 0. \quad (1.4)$$

První část funkce měří nestabilitu opakování sezónní složky, zatímco druhá část měří, zdali se sezónní faktory vysčítají na nulu. Jestliže je hodnota funkce  $g(\mathbf{s})$  malá, potom naše sezónní složka splňuje velmi dobře obě naše podmínky.

Jako poslední požadujeme, aby součet trendové a sezónní složky co nejlépe aproximoval naši časovou řadu, tj. snažíme se minimalizovat chybovou složku  $\boldsymbol{\epsilon}$ . Definujme funkci  $h(\mathbf{y} - \mathbf{z} - \mathbf{s}) : \mathbb{R}^T \rightarrow \mathbb{R}$

$$h(\mathbf{y} - \mathbf{z} - \mathbf{s}) := \sum_{t=1}^T (y_t - z_t - s_t)^2. \quad (1.5)$$

**Definice 18.** Mějme časovou řadu  $\{y_t\}_{t=1}^T$  s délkou opakování  $m$  a její aditivní dekompozici. Dále mějme zadané funkce  $f, g$  a  $h$  dle (1.3), (1.4) a (1.5). Potom definujeme

$$V(\mathbf{z}, \mathbf{s} | \mathbf{y}) := f(\mathbf{z}) + g(\mathbf{s}) + h(\mathbf{y} - \mathbf{z} - \mathbf{s}).$$

Trendovou a sezónní složku časové řady  $\{y_t\}_{t=1}^T$  dostaneme minimalizací této funkce vzhledem k  $\mathbf{z}$  a  $\mathbf{s}$ .

**Věta 2.** Mějme funkce  $f, g$  a  $h$  definované dle (1.3), (1.4) a (1.5) a časovou řadu  $\{y_t\}_{t=1}^T$  s délkou opakování  $m$ . Potom existuje právě jedno řešení minimalizačního problému z definice 18 dle rovnice:

$$\mathbf{H} \cdot \begin{pmatrix} \mathbf{z} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{y} \end{pmatrix},$$

kde

$$\mathbf{H} := \begin{pmatrix} \alpha \mathbf{P}^\top \mathbf{P} + \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \beta \mathbf{Q}^\top \mathbf{Q} + \gamma \mathbf{R}^\top \mathbf{R} + \mathbf{I} \end{pmatrix}$$

velikosti  $(2T) \times (2T)$ , kde  $\mathbf{I}$  je identická matice velikosti  $T \times T$ ,

$$\mathbf{P} := \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

velikosti  $(T-2) \times T$ ,

$$\mathbf{Q} := \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & -1 \end{pmatrix}$$

sloupec  $m+1$   
↓

velikosti  $(T-m) \times T$  a

$$\mathbf{R} := \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

sloupec  $m$   
↓

velikosti  $(T-m+1) \times T$ .



*Důkaz.* Jednotlivé kroky důkazu jsou podrobně popsány ve článku Schlicht (1981, str.375–376). □

Abychom mohli na naši časovou řadu aplikovat tuto větu, musíme ještě vyřešit problém výběru konstant  $\alpha$ ,  $\beta$  a  $\gamma$ . Konstanty můžeme zvolit dle libosti. Pro větší hodnoty  $\alpha$  dostáváme hladší trendovou složku a pro větší  $\beta$  a  $\gamma$  získáme stabilnější sezónní opakování, avšak samozřejmě za cenu zvětšení chybové složky.

Alternativně můžeme na volbu parametrů znovu nahlížet jako na optimalizační problém. Například budeme chtít zvolit parametry tak, aby chybová složka projevovала pouze slabé sezónní opakování. Mnohem zajímavější je však příklad, kde bychom chtěli minimalizovat tzv. ex post predikční chybu během posledních pár pozorování (například  $2m$ ) zvolením vhodných vah. Ex post predikční chybou máme na mysli rozdíl  $s_t(\alpha, \beta, \gamma, T)$  a  $s_t(\alpha, \beta, \gamma, t)$ , kde  $s_t(\alpha, \beta, \gamma, t)$  je hodnota sezónní složky v čase  $t$  vypočítaná pomocí Schlichtovy metody s parametry  $\alpha$ ,  $\beta$  a  $\gamma$  a s použitím pouze prvních  $t < T$  pozorování. Podobně můžeme tuto chybu vypočítat i pro trendovou složku.

Je zřejmé, že zvolení vhodných parametrů závisí na tvaru naší časové řady. Pokud má být sezónní opakování velmi konzistentní po celou délku řady je vhodné zvolit velké  $\beta$  a  $\gamma$ . Pokud však velikost sezónní složky závisí na velikosti trendové složky, potom je vhodné použít naopak malé hodnoty pro  $\beta$  a  $\gamma$ .

## 1.6 ETS metody

ETS metody neboli také metody exponenciálního vyrovnávání je skupina metod, které jsou řízené specifickými regresními rovnicemi. Celkem se jedná o patnáct metod a třicet modelů, které se liší právě rovnicemi jejich regresního modelu. Správnou metodu vybíráme podle toho, jakou charakteristiku vykazuje zadaná časová řada  $\{y_t\}_{t=1}^T$ . V této podkapitole budeme čerpat z knihy Forecasting with exponential smoothing: the state space approach (Hyndman a kol., 2008), článku z časopisu Journal of the American Statistical Association (Hyndman a Khandakar, 2008) a internetového zdroje Svetunkov (2022).

Jako první je třeba si představit obecný stavový model.

**Definice 19.** *Mějme časovou řadu  $\{y_t\}_{t=1}^T$  a  $\{\mathbf{x}_t\}_{t=1}^T$  posloupnost stavových vektorů, obsahující jednotlivé komponenty tvořící trendovou a sezónní složku časové řady. Poté obecný stavový model může být popsán rovnicemi:*

$$y_t = w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1})\epsilon_t, \quad (1.6)$$

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}) + \mathbf{g}(\mathbf{x}_{t-1})\epsilon_t, \quad (1.7)$$

kde  $\{\epsilon_t\}_{t=1}^T$  je chybová složka časové řady, funkce  $w$  a  $r$  jsou skalární pole a funkce  $\mathbf{f}$  a  $\mathbf{g}$  jsou vektorová pole.

Rovnice 1.6 bývá označována jako rovnice měření. Určuje vztah mezi vektorem stavů  $\mathbf{x}_{t-1}$  a napozorovanou hodnotou  $y_t$ . A rovnice 1.7, označována jako rovnice přechodu, popisuje vývoj vektoru stavu v čase. Pro různé volby funkcí obecného stavového modelu dostaneme jednotlivé ETS metody.

Jak jsme již zmínily, tak výběr jednotlivé metody záleží na charakteru dané časové řady. Jako první začneme s trendovou složkou  $\{z_t\}_{t=1}^T$ , která se skládá ze dvou částí, a to hladiny  $\{\ell_t\}_{t=0}^{T-1}$  a růstu  $\{b_t\}_{t=0}^{T-1}$ . Tyto dvě složky můžeme zkombinovat mnoha způsoby, z nichž si pět představíme:

$$\begin{aligned} \text{Žádná (N):} \quad z_t &= \ell_{t-1}, \\ \text{Aditivní (A):} \quad z_t &= \ell_{t-1} + b_{t-1}, \\ \text{Tlumená aditivní (A}_d\text{):} \quad z_t &= \ell_{t-1} + \phi b_{t-1}, \\ \text{Multiplikativní (M):} \quad z_t &= \ell_{t-1} \cdot b_{t-1}, \\ \text{Tlumená multiplikativní (M}_d\text{):} \quad z_t &= \ell_{t-1} \cdot b_{t-1}^\phi, \end{aligned}$$

kde  $\phi$  je tlumící parametr mezi 0 a 1.

Taktéž sezónní složka  $\{s_t\}_{t=0}^{T-1}$  může být různého typu, buďto se nemusí v časové řadě vůbec vyskytovat (označujeme N), nebo může být aditivní (A), či multiplikativní (M) vzhledem k trendové složce. A v poslední řadě složka chybová  $\{\epsilon_t\}_{t=1}^T$  může být pouze aditivní (A) či multiplikativní (M) vzhledem k trendu a sezónnosti. Pro chybovou složku platí, že jsou to jednotlivé realizace navzájem nezávislých náhodných veličin  $E_t$ , které mají normální rozdělení se střední hodnotou 0 a rozptylem  $\sigma^2$  (označujeme  $E_t \sim \text{NID}(0, \sigma^2)$ ).

Pokud zkombinujeme všechny možnosti pro růstovou složku se všemi možnostmi pro sezónní složku, dostaneme 15 možností nebo-li metod, které můžeme vidět v tabulce 1.2. Pro nás budou samozřejmě relevantní pouze kombinace při nichž se bude sezónnost vyskytovat.

		Sezónní složka		
		N	A	M
	Trendová složka	Žádná	Aditivní	Multiplikativní
N	Žádná	N,N	N,A	N,M
A	Aditivní	A,N	A,A	A,M
A <sub>d</sub>	Tlumená aditivní	A <sub>d</sub> ,N	A <sub>d</sub> ,A	A <sub>d</sub> ,M
M	Multiplikativní	M,N	M,A	M,M
M <sub>d</sub>	Tlumená multiplikativní	M <sub>d</sub> ,N	M <sub>d</sub> ,A	M <sub>d</sub> ,M

Tabulka 1.2: Možné kombinace povah růstové a sezónní složky

Některé dvojice jsou známé pod jinými názvy. Například (N,N) popisuje jednoduché exponenciální vyrovnávání (SES), dvojice (A,N) je známá pod jménem lineární Holtova metoda. A dvojice (A,A) a (A,M) popisují již dříve popsanou aditivní a multiplikativní Holtovu-Wintersovu metodu.

Když k této dvojici ještě přidáme buďto aditivní, či multiplikativní chybovou složku, dostaneme model ETS (error, trend, seasonality). Například ETS(A,A,M) označuje model, který uvažuje řadu, která je složena z aditivní složky chybové a růstové a má multiplikativní sezónní složku. Označení ETS( $\cdot, \cdot, \cdot$ ) nám pomáhá si vybavit pořadí, ve kterém jsou dané složky specifikovány.

U exponenciálního vyrovnávání si musíme dávat pozor na rozdíl mezi metodou a samotným modelem. Metoda exponenciálního vyrovnávání je pouze algoritmus,

díky kterému můžeme vytvářet predikce naší časové řadě. Avšak model je stochastický stavový prostor, který dává stejné predikce, avšak poskytuje i statistický rámec pro tvorbu predikčních intervalů a dalších vlastností.

Tabulky 1.4 a 1.5 na straně 17 a 18 popisují soustavy regresních rovnic všech ETS modelů, kde  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$  a  $0 < \phi < 1$  jsou konstanty,  $m$  je délka opakování sezónní složky a rovnice platí pro všechna  $t = 1, 2, \dots, T$ .

Jelikož předpokládáme normalitu chybové složky, můžeme model odhadnout pomocí maximální věrohodnosti. Dále, abychom mohli použít ETS metodu, potřebujeme zjistit počáteční hodnotu  $\mathbf{x}_0$ . Běžně se  $\mathbf{x}_0$  odhaduje společně s vyrovnávacími parametry  $\alpha$ ,  $\beta$ ,  $\gamma$  a  $\phi$  během procesu maximalizace věrohodnosti. Dostáváme tedy optimalizační úlohu, která může obsahovat velké množství parametrů. Navíc rozptyl chybové složky je taktéž považován za parametr, a tedy celkový počet parametrů v různých modelech je daný tabulkou 1.3.

Trend	Sezónnost	Počet parametrů	Odhady
N	N	3	$\hat{\ell}_0, \hat{\alpha}$ a $\hat{\sigma}^2$
A / M	N	5	$\hat{\ell}_0, \hat{b}_0, \hat{\alpha}, \hat{\beta}$ a $\hat{\sigma}^2$
A <sub>d</sub> / M <sub>d</sub>	N	6	$\hat{\ell}_0, \hat{b}_0, \hat{\alpha}, \hat{\beta}, \hat{\phi}$ a $\hat{\sigma}^2$
N	A / M	4+m-1	$\hat{\ell}_0, \hat{s}_{-m+2}, \hat{s}_{-m+3}, \dots, \hat{s}_0, \hat{\alpha}, \hat{\gamma}$ a $\hat{\sigma}^2$
A / M	A / M	6+m-1	$\hat{\ell}_0, \hat{b}_0, \hat{s}_{-m+2}, \hat{s}_{-m+3}, \dots, \hat{s}_0, \hat{\alpha}, \hat{\beta}, \hat{\gamma}$ a $\hat{\sigma}^2$
A <sub>d</sub> / M <sub>d</sub>	A / M	7+m-1	$\hat{\ell}_0, \hat{b}_0, \hat{s}_{-m+2}, \hat{s}_{-m+3}, \dots, \hat{s}_0, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\phi}$ a $\hat{\sigma}^2$

Tabulka 1.3: Počet parametrů pro jednotlivé ETS metody

Abychom dokázali rozložit časovou řadu na jednotlivé složky a taktéž vytvořit její predikci pro následující období, musíme komponenty časové řady a vyrovnávací parametry nahradit jejími odhady (například  $\alpha$  odhadneme jako  $\hat{\alpha}$ ). Po odhadnutí všech parametrů našich modelů budeme moci vytvořit předpovědi pro následné chování časové řady. Jelikož střední hodnota  $\epsilon_t$  je 0, rekursivní vzorce pro časovou řadu i její predikce bude stejná, jak pro aditivní, tak i multiplikační povahu chybové složky. Toto je shrnuto v tabulce 1.6 na straně 19, kde  $\hat{y}_{t+h|t}$  značí predikci pro čas  $t+h$ , která je založena na prvních  $t$  pozorováních,  $\hat{\phi}_h = \hat{\phi} + \hat{\phi}^2 + \dots + \hat{\phi}^h$  a  $h_m^+ = ((h-1) \bmod m) + 1$ .

Nyní, když již umíme odhadnout parametry a vytvářet predikce pomocí ETS metod, je důležité si vybrat, kterou metodu použijeme. Výběr zvolíme pomocí Akaikeova informačního kritéria:

$$\text{AIC} = -2 \log \hat{L} + 2k,$$

kde  $\hat{L}$  je věrohodnost modelu, která závisí na odhadech  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\phi})^\top$  a  $\hat{x}_0$  a  $k$  je počet odhadnutých parametrů.

Kritérium AIC můžeme upravit vzhledem ke zkreslení vytvořené použitím malého množství dat pomocí AIC<sub>c</sub>:

$$\text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{T-k-1},$$

nebo pomocí Bayesovského informačního kritéria:

$$\text{BIC} = \text{AIC} + k(\log T - 2).$$

Zvolíme metodu, která má minimální hodnotu informačního kritéria dle zvoleného kritéria.

Růst	Sezónnost		
	N	A	M
N	$y_t = \ell_{t-1} + \epsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \epsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \epsilon_t$
	$\ell_t = \ell_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1} + \alpha\frac{\epsilon_t}{s_{t-m}}$
		$s_t = s_{t-m} + \gamma\epsilon_t$	$s_t = s_{t-m} + \gamma\frac{\epsilon_t}{\ell_{t-1}}$
A	$y_t = \ell_{t-1} + b_{t-1} + \epsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \epsilon_t$
	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\frac{\epsilon_t}{s_{t-m}}$
	$b_t = b_{t-1} + \beta\epsilon_t$	$b_t = b_{t-1} + \beta\epsilon_t$	$b_t = b_{t-1} + \beta\frac{\epsilon_t}{s_{t-m}}$
		$s_t = s_{t-m} + \gamma\epsilon_t$	$s_t = s_{t-m} + \gamma\frac{\epsilon_t}{\ell_{t-1} + b_{t-1}}$
Ad	$y_t = \ell_{t-1} + \phi b_{t-1} + \epsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \epsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \epsilon_t$
	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\frac{\epsilon_t}{s_{t-m}}$
	$b_t = \phi b_{t-1} + \beta\epsilon_t$	$b_t = \phi b_{t-1} + \beta\epsilon_t$	$b_t = \phi b_{t-1} + \beta\frac{\epsilon_t}{s_{t-m}}$
		$s_t = s_{t-m} + \gamma\epsilon_t$	$s_t = s_{t-m} + \gamma\frac{\epsilon_t}{\ell_{t-1} + \phi b_{t-1}}$
M	$y_t = \ell_{t-1}b_{t-1} + \epsilon_t$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \epsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \epsilon_t$
	$\ell_t = \ell_{t-1}b_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1}b_{t-1} + \alpha\epsilon_t$	$\ell_t = \ell_{t-1}b_{t-1} + \alpha\frac{\epsilon_t}{s_{t-m}}$
	$b_t = b_{t-1} + \beta\frac{\epsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1} + \beta\frac{\epsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1} + \beta\frac{\epsilon_t}{\ell_{t-1}s_{t-m}}$
		$s_t = s_{t-m} + \gamma\epsilon_t$	$s_t = s_{t-m} + \gamma\frac{\epsilon_t}{\ell_{t-1}b_{t-1}}$
Md	$y_t = \ell_{t-1}b_{t-1}^\phi + \epsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m} + \epsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m} + \epsilon_t$
	$\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\epsilon_t$	$\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\epsilon_t$	$\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\frac{\epsilon_t}{s_{t-m}}$
	$b_t = b_{t-1}^\phi + \beta\frac{\epsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1}^\phi + \beta\frac{\epsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1}^\phi + \beta\frac{\epsilon_t}{\ell_{t-1}s_{t-m}}$
		$s_t = s_{t-m} + \gamma\epsilon_t$	$s_t = s_{t-m} + \gamma\frac{\epsilon_t}{\ell_{t-1}b_{t-1}}$

Tabulka 1.4: Popis stavového prostoru ETS modelů s aditivní chybovou složkou

Růst	Sezónnost	
	N	A
		M
N	$y_t = \ell_{t-1}(1 + \epsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\epsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \epsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\epsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\epsilon_t$
A	$y_t = (\ell_{t-1} + b_{t-1}) + (1 + \epsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) + (1 + \alpha\epsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\epsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m}) + (1 + \epsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\epsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\epsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\epsilon_t$
A <sub>d</sub>	$y_t = (\ell_{t-1} + \phi b_{t-1}) + (1 + \epsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) + (1 + \alpha\epsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\epsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \epsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\epsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\epsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\epsilon_t$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \epsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\epsilon_t)$ $b_t = b_{t-1}(1 + \beta\epsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \epsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\epsilon_t$ $b_t = b_{t-1} + \beta \frac{\ell_{t-1}b_{t-1} + s_{t-m}}{\ell_{t-1}}\epsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\epsilon_t$
M <sub>d</sub>	$y_t = \ell_{t-1}b_{t-1}^\phi(1 + \epsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\epsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\epsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \epsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\epsilon_t$ $b_t = b_{t-1}^\phi + \beta \frac{\ell_{t-1}b_{t-1}^\phi + s_{t-m}}{\ell_{t-1}}\epsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\epsilon_t$

Tabulka 1.5: Popis stavového prostoru ETS modelů s multiplikativní chybovou složkou

Růst	Sezónnost		
	N	A	M
N	$\hat{\ell}_t = \hat{\alpha}y_t + (1 - \hat{\alpha})\hat{\ell}_{t-1}$	$\hat{\ell}_t = \hat{\alpha}(y_t - \hat{s}_{t-m}) + (1 - \hat{\alpha})\hat{\ell}_{t-1}$	$\hat{\ell}_t = \hat{\alpha}(y_t/\hat{s}_{t-m}) + (1 - \hat{\alpha})\hat{\ell}_{t-1}$
	$\hat{s}_t = \hat{\gamma}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\gamma})\hat{s}_{t-m}$	$\hat{s}_t = \hat{\gamma}(y_t - \hat{\ell}_{t-1}) + (1 - \hat{\gamma})\hat{s}_{t-m}$	$\hat{s}_t = \hat{\gamma}(y_t/\hat{\ell}_{t-1}) + (1 - \hat{\gamma})\hat{s}_{t-m}$
	$\hat{y}_{t+h t} = \hat{\ell}_t$	$\hat{y}_{t+h t} = \hat{\ell}_t + \hat{s}_{t-m+h_m^+}$	$\hat{y}_{t+h t} = \hat{\ell}_t \hat{s}_{t-m+h_m^+}$
A	$\hat{\ell}_t = \hat{\alpha}y_t + (1 - \hat{\alpha})(\hat{\ell}_{t-1} + \hat{b}_{t-1})$	$\hat{\ell}_t = \hat{\alpha}(y_t - \hat{s}_{t-m}) + (1 - \hat{\alpha})(\hat{\ell}_{t-1} + \hat{b}_{t-1})$	$\hat{\ell}_t = \hat{\alpha}(y_t/\hat{s}_{t-m}) + (1 - \hat{\alpha})(\hat{\ell}_{t-1} + \hat{b}_{t-1})$
	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}$
	$\hat{y}_{t+h t} = \hat{\ell}_t + h\hat{b}_t$	$\hat{y}_{t+h t} = \hat{\ell}_t + h\hat{b}_t + \hat{s}_{t-m+h_m^+}$	$\hat{y}_{t+h t} = (\hat{\ell}_t + h\hat{b}_t)\hat{s}_{t-m+h_m^+}$
A <sub>d</sub>	$\hat{\ell}_t = \hat{\alpha}y_t + (1 - \hat{\alpha})(\hat{\ell}_{t-1} + \hat{\phi}\hat{b}_{t-1})$	$\hat{\ell}_t = \hat{\alpha}(y_t - \hat{s}_{t-m}) + (1 - \hat{\alpha})(\hat{\ell}_{t-1} + \hat{\phi}\hat{b}_{t-1})$	$\hat{\ell}_t = \hat{\alpha}(y_t/\hat{s}_{t-m}) + (1 - \hat{\alpha})(\hat{\ell}_{t-1} + \hat{\phi}\hat{b}_{t-1})$
	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{\phi}\hat{b}_{t-1}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{\phi}h\hat{b}_{t-1}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t - \hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{\phi}\hat{b}_{t-1}$
	$\hat{y}_{t+h t} = \hat{\ell}_t + \hat{\phi}_h\hat{b}_t$	$\hat{y}_{t+h t} = \hat{\ell}_t - \hat{\ell}_{t-1} - \hat{\phi}\hat{b}_{t-1} + (1 - \hat{\gamma})\hat{s}_{t-m}$	$\hat{y}_{t+h t} = \hat{\ell}_t + \hat{\phi}_h\hat{b}_t + \hat{s}_{t-m+h_m^+}$
M	$\hat{\ell}_t = \hat{\alpha}y_t + (1 - \hat{\alpha})\hat{\ell}_{t-1}\hat{b}_{t-1}$	$\hat{\ell}_t = \hat{\alpha}(y_t - \hat{s}_{t-m}) + (1 - \hat{\alpha})\hat{\ell}_{t-1}\hat{b}_{t-1}$	$\hat{\ell}_t = \hat{\alpha}(y_t/\hat{s}_{t-m}) + (1 - \hat{\alpha})\hat{\ell}_{t-1}\hat{b}_{t-1}$
	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t/\hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t/\hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t/\hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}$
	$\hat{y}_{t+h t} = \hat{\ell}_t\hat{b}_t^h$	$\hat{y}_{t+h t} = \hat{\ell}_t + \hat{\phi}_h\hat{b}_t + \hat{s}_{t-m+h_m^+}$	$\hat{y}_{t+h t} = (\hat{\ell}_t + \hat{\phi}_h\hat{b}_t)\hat{s}_{t-m+h_m^+}$
M <sub>d</sub>	$\hat{\ell}_t = \hat{\alpha}y_t + (1 - \hat{\alpha})\hat{\ell}_{t-1}\hat{b}_{t-1}^{\hat{\phi}}$	$\hat{\ell}_t = \hat{\alpha}(y_t - \hat{s}_{t-m}) + (1 - \hat{\alpha})\hat{\ell}_{t-1}\hat{b}_{t-1}^{\hat{\phi}}$	$\hat{\ell}_t = \hat{\alpha}(y_t/\hat{s}_{t-m}) + (1 - \hat{\alpha})\hat{\ell}_{t-1}\hat{b}_{t-1}^{\hat{\phi}}$
	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t/\hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}^{\hat{\phi}}$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t/\hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}^{\hat{\phi}}hi$	$\hat{b}_t = \hat{\beta}(\hat{\ell}_t/\hat{\ell}_{t-1}) + (1 - \hat{\beta})\hat{b}_{t-1}^{\hat{\phi}}hi$
	$\hat{y}_{t+h t} = \hat{\ell}_t\hat{b}_t^{\hat{\phi}_h}$	$\hat{y}_{t+h t} = \hat{\ell}_t\hat{b}_t^{\hat{\phi}_h} + \hat{s}_{t-m+h_m^+}$	$\hat{y}_{t+h t} = \hat{\ell}_t\hat{b}_t^{\hat{\phi}_h}\hat{s}_{t-m+h_m^+}$

Tabulka 1.6: Vzorce pro rekurzivní kalkulace a predikce ETS modelů

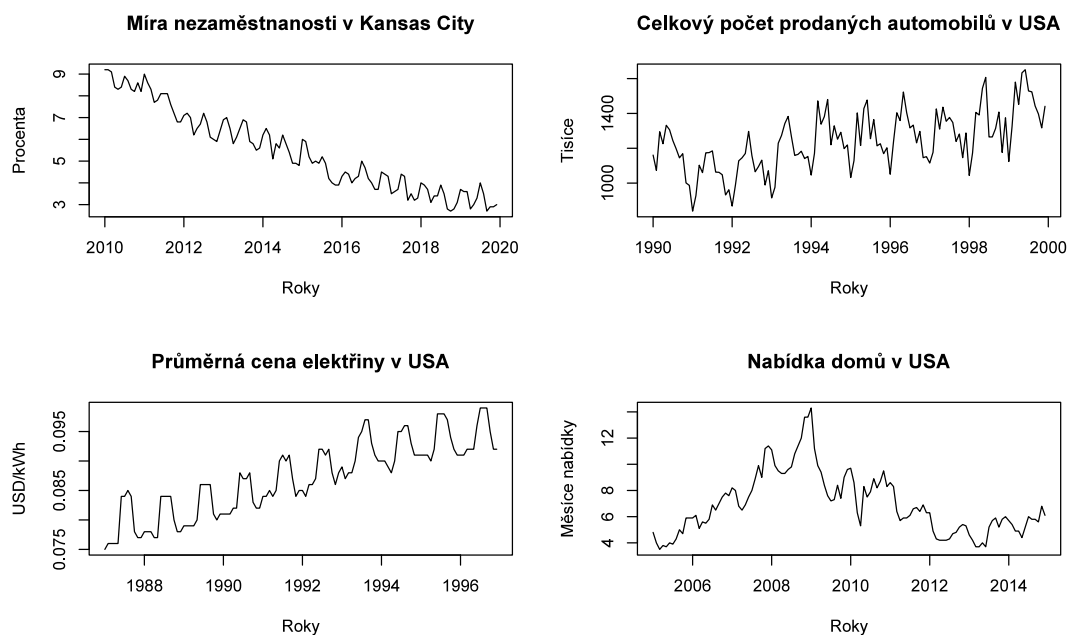
## 2. Praktická část

V druhé části této bakalářské práce se budeme věnovat aplikacím všech metod popsaných v teoretické části. Metody aplikujeme na reálné ekonomické časové řady a výsledky opatříme komentářem, který podpoříme graficky i numericky. Následně přejdeme ke komparaci metod dle jejich přesnosti predikce.

### 2.1 Data

Reálná data, které budeme v této části využívat, jsme získali z webových stránek Economic Research - Federal Reserve Bank of Saint Louis. Data jsou ve formě desetiletých časových řad s měsíční frekvencí, tedy celkem 120 pozorování. Ty vykazují sezónnost, takže jsou pro nás vhodné ke zkoumání. Z jednotlivých řad budeme vždy odebírat 20 posledních pozorování, a tedy prezentované metody aplikujeme pouze na prvních 100 pozorování. Poté jakmile budeme chtít porovnávat metody dle přesnosti predikce, využijeme těchto posledních 20 pozorování jakožto validační data.

Obrázek 2.1, tabulka 2.1 a tabulka 2.2 nám dává základní přehled o zkoumaných časových řadách.



Obrázek 2.1: Grafy jednotlivých časových řad



	Název	Jednotky	Časové období
(1)	Míra nezaměstnanosti v Kansas City	Procenta	1/2010-12/2019
(2)	Celkový počet prodaných automobilů v USA	Tisíce	1/1990-12/1999
(3)	Průměrná cena elektřiny v USA	USD/kWh	1/1987-12/1996
(4)	Nabídka domů v USA	Měsíce nabídky	1/2005-12/2014

Tabulka 2.1: Popis jednotlivých časových řad

	Průměr	Směrodatná odchylka	Průměrný absolutní přírůstek
(1)	5.45	1.85	0.34
(2)	1240	168	104
(3)	0.087	0.006	0.0014
(4)	6.93	2.35	0.59

Tabulka 2.2: Základní popisné charakteristiky jednotlivých časových řad

## 2.2 Software

K výpočtům budeme používat statistický software R. Navíc pro aplikaci Holtovy-Wintersovy metody a ETS metod budeme používat volně stáhnutelný balíček `forecast`. Tento balíček dokáže pro ETS modely pomocí funkce `ets()` automaticky vypočítat informační kritérium  $AIC_c$  a vybrat nejvhodnější model, stejně tak pomocí metody maximální věrohodnosti odhadnout počáteční hodnoty a vyhlazovací parametry. Taktéž budeme používat funkci `HoltWinters()` pro Holtovu-Wintersovu metodu, která dokáže pomocí vlastního heuristického algoritmu získat počáteční hodnoty a taktéž vyrovnávacích konstanty. Avšak, jelikož jsme si teoretické části ukázali, jak zvolíme počáteční hodnoty, bylo by vhodné tuto teorii využít i v praxi. Stejně tak jsme doporučili fixní volbu vyrovnávacích konstant  $\alpha$ ,  $\beta$  a  $\gamma$ .

## 2.3 Aplikace jednotlivých přístupů na reálná data

### 2.3.1 Jednoduché přístupy k sezónnosti

Přístupy, které budeme v této podkapitole používat, jsou jednoduché, jak z početního hlediska, tak i v počtu vstupů, které potřebuje, abychom je mohli aplikovat. Ve své podstatě potřebujeme pouze časovou řadu. Zde se nemusíme rozhodovat, jaké počáteční hodnoty či vyrovnávací konstanty použijeme. Tato jednoduchost však má svou cenu, a to je malá flexibilita. Pouze pro predikci můžeme volit tvar křivky, dle které budeme očistěnou časovou řadu vyrovnávat a následně predikovat její budoucí vývoj. Jelikož se jedná o reálné ekonomické časové řady, tak pro predikci trendu budeme vyrovnávat posledních 20 pozorování pomocí přímky. I když tato metoda nenabízí téměř žádnou flexibilitu, můžeme navzájem porovnat, která verze této metody je vhodnější, zdali aditivní či multiplikativní.

Jelikož jednotlivé metody používáme pouze na prvních 100 pozorování, můžeme poté porovnat jejich predikci s validačními daty, které jsme ve výpočtu

nepoužili. Existuje mnoho způsobů, jak lze toto porovnání provádět. My si zde představíme tři nejjednodušší způsoby, které byly představeny ve článku Watsona (Watson, 2012). Nejjednodušeji získáme míru přesnosti predikce pomocí Mean Absolute Error (MAE), která vypočítá průměrnou absolutní odchylku predikce od reality. Jedním z problémů MAE je, že relativní velikost chyby není intuitivně zřejmá a může být složité rozeznat velkou chybu od malé chyby. Abychom se zbavili tohoto problému, můžeme použít Mean Absolute Percentage Error (MAPE), kterou vypočítáme jako relativní velikost chyby vzhledem k hodnotě časové řady, a tedy nám udává absolutní chybu v procentech. Zde je nutné si říci, že nelze jednoduše porovnávat velikost chyby mezi různými časovými řadami a na jejím základě rozhodovat, kterou časovou řadu jsme odpredikovali nejlépe, protože zde hraje velkou roli relativní velikost sezónních výkyvů, volatilita, změny trendu a jiné.

Jelikož oba způsoby využívají pouze průměrnou chybu, mohou podceňovat dopad, sice málo častých, ale velkých výkyvů. Abychom se vypořádali s tímto problémem, použijeme Root Mean Square Error (RMSE), která umocněním chyby na druhou, zvyšuje váhu velkých chyb. Pokud porovnáme MAE s RMSE, zjistíme zdali predikce obsahuje, sice málo časté, ale velké chyby. Čím větší je rozdíl mezi MAE a RMSE, tím více je velikost chyby nekonzistentní. Jednotlivé způsoby pro porovnání predikce vypočítáme následovně:

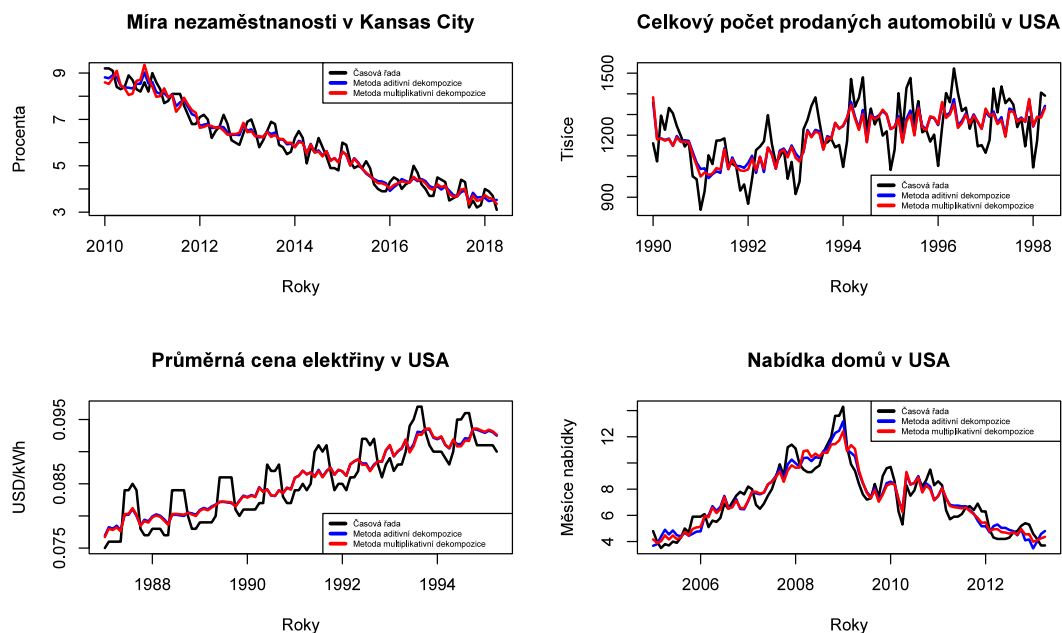
$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{h=1}^{h=n} |y_{T+h} - \hat{y}_{T+h|T}|, \\ \text{MAPE} &= \frac{100\%}{n} \sum_{h=1}^{h=n} \left| \frac{y_{T+h} - \hat{y}_{T+h|T}}{y_{T+h}} \right|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{h=1}^{h=n} (y_{T+h} - \hat{y}_{T+h|T})^2}, \end{aligned}$$

kde  $n$  je počet porovnávaných predikcí,  $T$  je počet pozorování, které jsme použili na vytvoření predikce,  $y_{T+h}$  je reálná hodnota v čase  $T+h$  a  $\hat{y}_{T+h|T}$  je predikce pro časovou řadu v čase  $T+h$  dle prvních  $T$  pozorování.

Nyní již můžeme přejít k aplikaci jednoduchých metod na naše časové řady a porovnat přesnost jejich predikce. Výsledky jsou shrnuty v obrázku 2.2, obrázku 2.3 a tabulce 2.3.

Číslo řady	Metoda	MAE	MAPE	RMSE
(1)	Aditivní	0.33	10.5 %	0.40
	Multiplikativní	0.33	9.9 %	0.40
(2)	Aditivní	108	7.4 %	119
	Multiplikativní	107	7.4 %	118
(3)	Aditivní	0.0012	1.2 %	0.0013
	Multiplikativní	0.0010	1.1 %	0.0012
(4)	Aditivní	2.36	43.3 %	2.53
	Multiplikativní	2.41	43.2 %	2.59

Tabulka 2.3: MAE, MAPE a RMSE jednoduchých přístupů pro jednotlivé časové řady

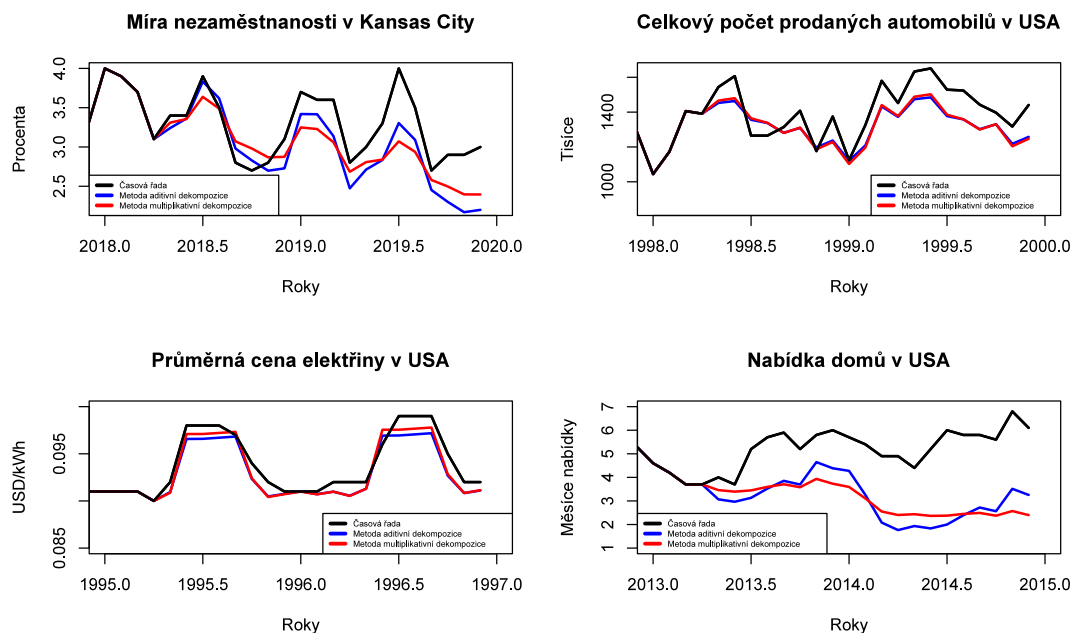


Obrázek 2.2: Časové řady očištěné od sezónní složky dle jednoduchých přístupů

Mezi aditivní a multiplikační metodou nepozorujeme žádné výrazné rozdíly, co se týče velikostí chyb. To, že u některých časových řad jsme si vedli lépe než u jiných, můžeme vidět, jednak z relativní velikosti chyb, ale také v obrázku 2.3. Nejlépe jsme si vedli v predikci pro časovou řadu (3), kde jsme dokázali velmi dobře predikovat jak sezónní, tak i trendovou složku. Když se podíváme na časovou řadu (1), tak můžeme vidět, že jsme se po prvních 6 pozorováních odchýlili od reality, což je nejspíše dáno změnou trendu dané časové řady. Avšak vidíme, že reálné sezónní vychýlení je podobné sezónnímu vychýlení pro aditivní metodu. Naopak multiplikační metoda velikost sezónní složky výrazně podceňuje. Lze tedy vyvodit závěr, že je lepší na tuto časovou řadu využít aditivní přístup, pokud bychom chtěli tuto časovou řadu predikovat v budoucnu.

Nyní se podívejme na časovou řadu (2), kde můžeme vidět, že pro prvních asi 10 pozorování jsme velmi dobře dokázali predikovat trendovou složku, avšak nedokázali predikovat výkyvy reálné časové řady. Naopak u posledních 10 pozorování vidíme, že jsme dobře predikovali sezónní vychýlení, jak pro aditivní, tak i pro multiplikační verzi. Nicméně naše metody nebyly schopné dobře predikovat trendovou složku, což zapříčinilo, že naše predikce byly asi 150 tisíc menší, než realita. Důvodem je jednorázový nárůst na začátku roku 1999.

U časové řady (4) si můžeme všimnout, že jsme vůbec nedokázali odhadnout následné chování trendové složky, což se odrazilo na velikosti průměrné relativní absolutní chyby, která přesáhla 40 % u obou metod. To je zcela jistě zapříčiněno změnou trendu, kterou jsme pomocí této metody nedokázali předpovídat. Také můžeme vidět nějaké náznaky podobných sezónních výkyvů mezi aditivní metodou a realitou. To samé nelze říct pro multiplikační metodu, která pro malé hodnoty má jen velmi malé výkyvy, což je však dáno tím, že hodnoty trendu jsou více než dvakrát menší, a tedy i sezónní vychýlení je více než dvakrát menší.



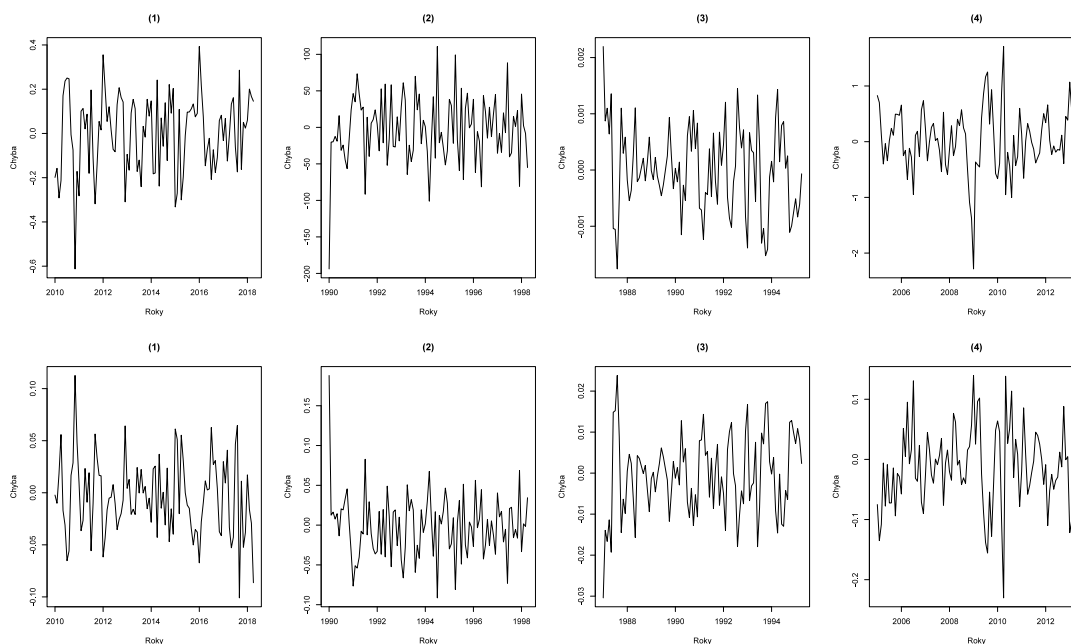
Obrázek 2.3: Grafy predikcí dle jednoduchých přístupů a reality pro jednotlivé časové řady

Rozdíly mezi MAE a RMSE u jednotlivých časových řad nám říkají, že zde nejsou žádné velké jednorázové výkyvy, takže predikce jsou velice konzistentní.

Ještě se můžeme podívat na chybovou složku a analyzovat ji. Mělo by platit, že to jsou realizace nezávislých náhodných veličin se střední hodnotou 0 a s konstantním rozptylem. Jelikož v teorii vytvoříme řadu očištěnou od trendové složky a řadu očištěnou od sezónní složky, jejím sečtením dostaneme časovou řadu, která je složena s trendové, sezónní a dvou chybových složek, jelikož ta je přítomna v obou očištěných řadách. Následným odečtením původní časové řady dostaneme samotnou chybovou složku pro aditivní metodu. Abychom ji dostali pro multiplikační dekompozici musíme součty a rozdíly nahradit součiny a podíly a nakonec ještě odečíst od celé řady jedničku.

Výsledek můžeme vidět v obrázku 2.4. V něm si lze všimnout a následně vypočítat, že průměr chybové složky je relativně blízký nule a taktéž rozptyl se zdá být celkem konstantní po celou dobu časové řady. Co však můžeme pozorovat je, že jednotlivé členy chybové složky na sobě nejsou zcela nezávislé. Toto můžeme vidět například u poslední časové řady pro obě metody, kde dvě po sobě následující realizace mají často podobnou hodnotu. I tak však můžeme být celkem spokojeni s vlastnostmi chybové složky, když uvážíme, že se jedná pouze o jednoduchou metodu.

Celkově však lze i tyto jednoduché metody označit za velice přesné a lze je doporučit pro predikci následného vývoje časové řady. Jako velké pozitivum musíme vyzdvihnout jednoduchost jejího použití i bez pomoci jakéhokoliv softwaru. Na druhou stranu mezi nedostatky této metody je jednoznačně již zmíněná neexistující flexibilita. A to v mnoha směrech. Tato metoda vůbec nepočítá s jakýmkoliv vývojem sezónních faktorů, což může vyústit v její nefunkčnost. Také



Obrázek 2.4: Chybová složka pro jednotlivé řady dle přístupu jednoduché aditivní a multiplikativní dekompozice

teorii pro následnou predikci jsme museli vytvořit navíc. Ta však dokáže pracovat pouze s trendem, který si sami zvolíme, což může být další komplikace.

### 2.3.2 Sezónnost modelována pomocí kvalitativní proměnné

Tato metoda je z dalších, které jsou ve své podstatě velice jednoduché a pro svůj chod nepotřebují žádné vstupní hodnoty ani vyhlazovací parametry. Zde můžeme vybírat pouze tvar trendové složky pomocí volby  $f(t)$ . Budeme se tedy zabývat, jak můžeme zlepšit predikci pomocí různých voleb funkce  $f(t)$ . Pro první tři časové řady jako první případ zvolíme  $f(t) = t$ . Je bezpředmětné tuto možnost zvolit pro poslední časovou řadu, jelikož je patrné, že trend není lineární. Další funkci, kterou můžeme dosadit je  $f(t) = 1 - 0.99^t$ . Zde jsme se inspirovali kapitolou o ETS metodách, přesněji možnosti existence tlumené verze trendové složky. Tato funkce je speciálně nadesignovaná pro časovou řadu se 100 pozorováními. Pokud bychom chtěli aplikovat podobnou funkci na časovou řadu s odlišným počtem pozorování, tak musíme nahradit číslo 0.99 vhodnější hodnotou. Tato funkce se chová podobně jako přímka, ale postupně se zpomaluje její rychlost změny, a nakonec vyústí v konstantní trend. Tuto křivku, stejně jako přímku, nelze použít pro časovou řadu (4).

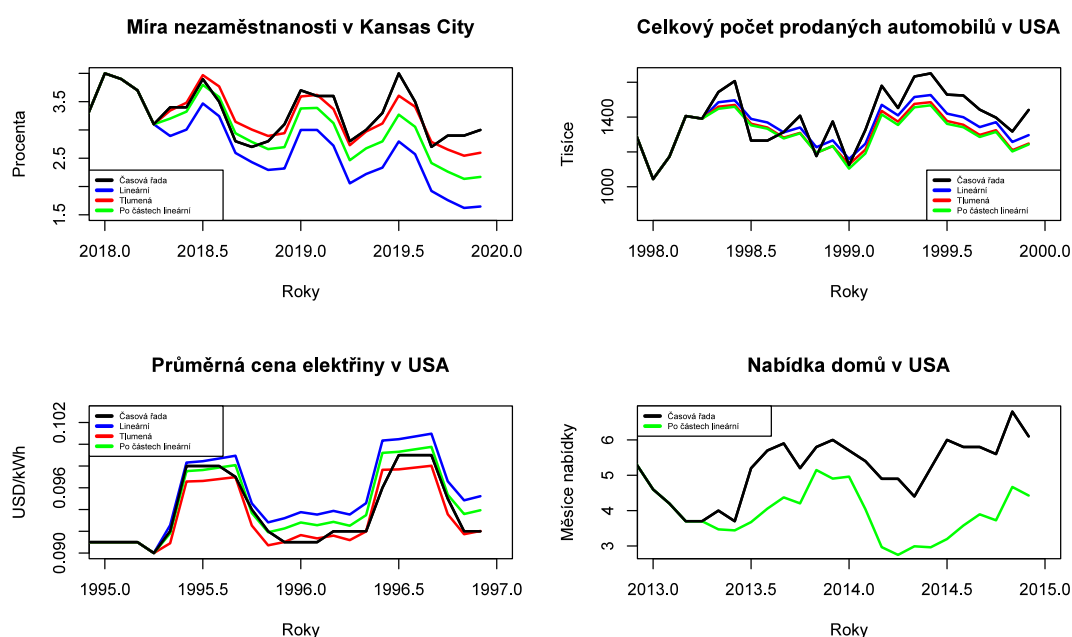
Funkci, kterou však můžeme použít na poslední řadu a taktéž na všechny předchozí, je po částech lineární křivka  $f(t) = t + (t - k_1)_+ + (t - k_2)_+ + \dots + (t - k_l)_+$ , kde  $(t - k_1)_+ = \max(0, t - k_1)$  a  $k_1$  až  $k_l$  jsou jednotlivé uzly, kde se mění lineární trend. Například právě pro řadu (4) můžeme vidět, že se trend mění na začátku roku 2009, na konci roku 2009, na konci roku 2010 a potom na začátku roku 2012, což by tedy znamenalo, že  $k_1 = 49$ ,  $k_2 = 55$ ,  $k_3 = 71$  a  $k_4 = 89$ . Pak by tedy trendová část regresní rovnice měla tvar  $\beta_0 + \beta_1 t + \beta_2 (t - 49)_+ + \beta_3 (t -$

$55)_+ + \beta_4(t - 71)_+ + \beta_5(t - 89)_+$ . Podobně zvolíme uzly pro ostatní časové řady dle tabulky 2.4.

Časová řada	(1)	(2)	(3)	(4)
$k_1$	13	13	76	49
$k_2$	25	55	-	55
$k_3$	67	-	-	71
$k_4$	-	-	-	89

Tabulka 2.4: Uzly pro jednotlivé časové řady

Nyní se již můžeme podívat, jak dobře tyto volby funkcí predikují následný vývoj jednotlivých časových řad v obrázku 2.5 a tabulce 2.5.



Obrázek 2.5: Grafy predikcí dle metody SMPKP a reality pro jednotlivé časové řady

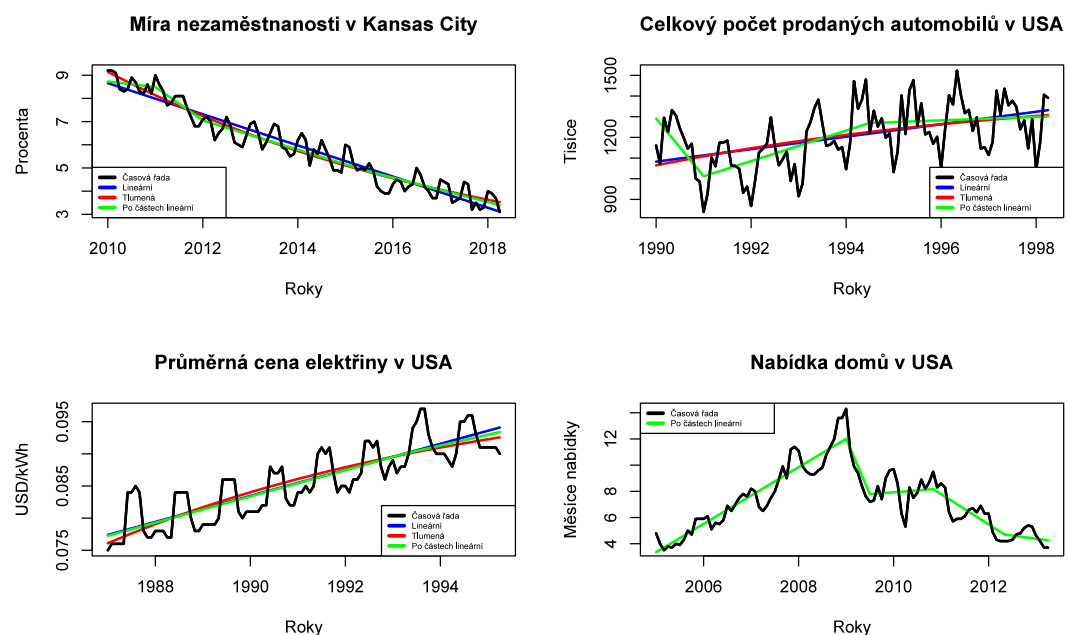
Jak lze očekávat, tak pro různé volby funkcí  $f(t)$  jsou výsledky dosti odlišené. Pro první a třetí časovou řadu se tlumená verze našeho modelu zdá nejlepší, navíc pokud to porovnáme s lineární verzí, tak je výrazně lepší. Je tedy možné brát dosazení  $1 - 0.99^t$  do funkce  $f$  jako velmi dobrou shodu pro všechny tři řady a celkově jako skvělou náhradu pro funkci  $x$ , když se potýkáme s řadou, která má celkem lineární trend, ale ke konci začíná trend trochu zpomalovat. Když se podíváme na obrázek 2.6, vidíme, že tato skutečnost je pravda pro první i třetí časovou řadu.

Pokud se podíváme na první časovou řadu, tak můžeme spatřit, že pro asi prvních 10 pozorování se tlumená verze a po částech lineární verze příliš neliší. Obě zde taktéž velmi dobře predikují vývoj. Nicméně v druhé části naší predikce si můžeme všimnout, že po částech lineární verze predikovala nižší hodnoty než

Číslo řady	Metoda	MAE	MAPE	RMSE
(1)	Lineární	0.74	23.1 %	0.81
	Tlumená	0.18	5.7 %	0.22
	Po částech lineární	0.35	11.1 %	0.43
(2)	Lineární	85	5.9 %	93
	Tlumená	109	7.5 %	121
	Po částech lineární	119	8.1 %	130
(3)	Lineární	0.0018	1.9 %	0.0021
	Tlumená	0.0008	0.9 %	0.0010
	Po částech lineární	0.0009	1.0 %	0.0012
(4)	Po částech lineární	1.53	28.0 %	1.66

Tabulka 2.5: MAE, MAPE a RMSE přístupů SMPKP pro jednotlivé časové řady

reálně nastaly. Naopak tlumená verze si udržovala celkem přesnou predikci po celé období predikce. Co se týče lineární verze, tak ta nedokázala předpovědět trendovou složku, což je dáno tím, že trend této časové řady není jednoduše lineární, ale mírně se mění a pokles zpomaluje. Tato skutečnost lze dobře pozorovat v obrázku 2.6.



Obrázek 2.6: Grafy trendové složky dle jednotlivých SMPKP přístupů pro jednotlivé časové řady

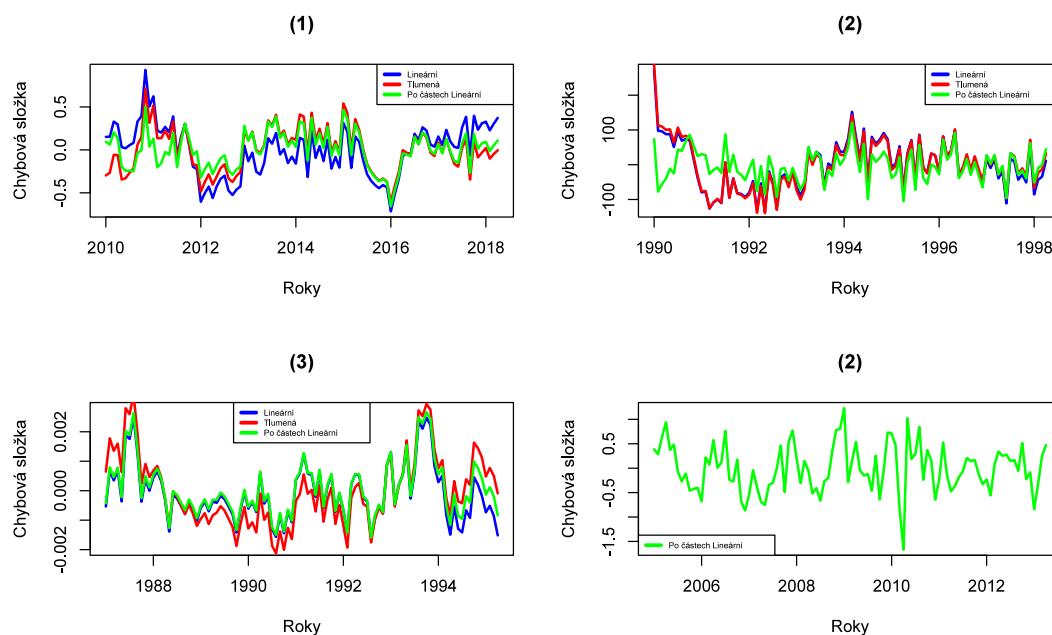
U druhé časové řady můžeme vidět, že predikce pro všechny tři verze jsou velice podobné. Vychýlení v první části se neshoduje s predikcí a nárůst časové řady na začátku roku 1999 naše predikce taktéž nezachytila. Naopak sezónní vychýlení jsme v tomto období dokázali velice přesně předpovědět.

Predikce pro časovou řadu (3) můžeme vyhodnotit jako velmi přesné, a dokonce pro tlumenou verzi jsme dostali velikost MAPE pod jedno procento. Mů-

žeme si všimnout, že lineární verze predikuje mírně vyšší hodnoty než ostatní metody, a taktéž více než realita. Toto je znovu dáno tím, že trend není čistě lineární, ale rychlost změny trendové složky zpomaluje, což lze opět dobře vidět v obrázku 2.6. Zajímavé je se podívat na to, že sezónní výkyvy jsou větší než všechny metody předpovídaly. Tlumená verze velmi dobře předpověděla následný vývoj pro časovou řadu v období, kdy je sezónnost záporná. Naopak verze s po částech lineárním trendem lépe předpovídala vývoj v období, kdy je řada vychýlena kladnou sezónností. Toto nejspíše znamená, že se sezónní složka třetí časové řady postupně zvětšuje, a tedy by bylo lepší využít jakoukoliv metodu s multiplikativní dekompozicí.

Přesnost predikce pro časovou řadu (4) je velice špatná (MAPE = 28%), avšak je to jednoznačně lepší, než predikce z předchozí sekce. Tato metoda tedy daleko lépe zachytila změnu trendu, avšak ne zcela. Jistě pokud bychom vybrali jiné uzly, tak můžeme dostat lepší predikci, což jsme však dříve nemohli vědět. Pokud se podíváme na sezónní vychýlení, tak zde naše predikce není špatná a pokud bychom trend zachytili, tak by mohla predikce být velice přesná.

Jestliže svoji pozornost přesuneme na konzistenci predikcí, tak můžeme vidět, že rozdíly mezi MAE a RMSE nejsou nikterak velké, ale v předchozí části tyto rozdíly byly menší, a tedy konzistence je trochu menší, avšak pořád velice přijatelná.



Obrázek 2.7: Chybová složka pro jednotlivé řady dle modelu SMPKP

Jako poslední se ještě podíváme na chybovou složku našich časových řad (obrázek 2.7). Průměr chybové složky je u všech čtyř řad pro všechny metody relativně blízký nule, což je přesně to, co požadujeme. Rozptyl se taktéž nezdá příliš proměnlivý pro různé části časových řad, i když u třetí časové řady vidíme celkem malý rozptyl mezi léty 1988 a 1990. Znovu však, stejně jako to bylo u jednoduchých metod, můžeme vidět, že je porušena nezávislost. Toto lze dobře



pozorovat na první a třetí časové řadě, kdy jednotlivé položky chybové složky těsně následující mají často stejný směr a velikost. Pro jednotlivé verze SMPKP metody je chybová složka velice podobná. Avšak i zde musíme být celkem spokojeni s charakterem chybové složky uvědomíme-li si, jak jednoduchou metodu používáme.

Modelování sezónnosti pomocí kvalitativní proměnné se ukazuje jako velmi dobrá metoda. Velkou výhodou této metody je, že lze volně volit tvar trendové složky. Toto může velmi dobře vylepšit predikci následného vývoje časové řady. Tento efekt lze dobře ilustrovat na volbě  $1 - 0.99^t$ . Přineslo nám to výrazně lepší predikci než lineární metoda. Taktéž tímto způsobem můžeme predikovat i řady, kde se trend mění drasticky, jak můžeme sledovat u poslední časové řady. Hlavní nevýhodou je nulová flexibilita u sezónní složky a dokonce i neexistence multiplikativní verze.

### 2.3.3 Holtova-Wintersova metoda

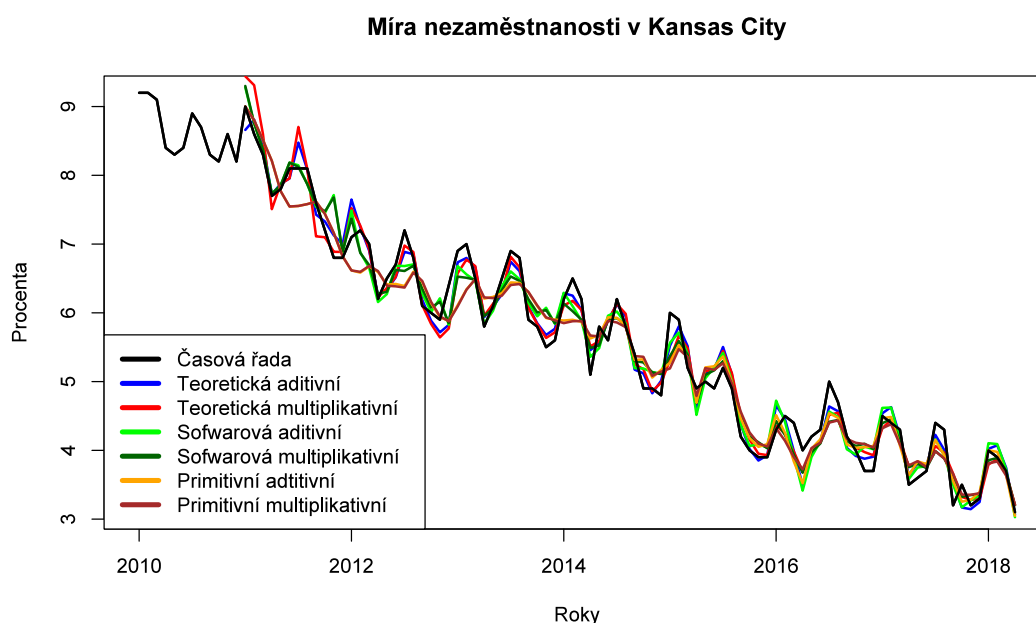
Holtova-Wintersova je první z metod, kde můžeme vybírat různé kombinace vyhlazovacích parametrů  $\alpha$ ,  $\beta$  a  $\gamma$  a počátečních hodnot  $\ell_0$ ,  $b_0$ ,  $s_{-11}$ ,  $s_{-10}$  až  $s_0$ . Tato volnost nám dává velmi dobrou flexibilitu, avšak musíme dávat pozor na to, jak tyto hodnoty vybereme, jelikož sice nám mohou velmi dobře posloužit, ale pokud je vybereme špatně, tak může nastat problém, který nám dá velmi špatnou predikci. Jako takovou bezpečnou variantu jsme již dříve doporučili  $\alpha = 0.4$ ,  $\beta = 0.1$  a  $\gamma = 0.4$ . Zároveň jsme si již ukázali, jak zvolíme jednotlivé počáteční hodnoty v teoretické části pro obě verze této metody. Jelikož v softwaru R pomocí funkce `HoltWinters()` můžeme zadávat tyto hodnoty přímo nebo nechat na softwaru, aby je vypočítal dle svého algoritmu, je lákavé tyto volby parametrů, které předkládáme, podrobit ověření.

#### Počáteční hodnoty

Prvně se podívejme na počáteční hodnoty. Vyhlazovací parametry zvolíme fixně dle našeho doporučení. Obě verze pro výběr počátečních hodnot porovnáme s algoritmem, který nám nabízí software a který by měl být velmi dobrou metodou. Z druhého konce spektra vezmeme jako porovnání velmi primitivní volbu počátečních hodnot a to  $\ell_0 = y_1$ ,  $b_0 = y_{13} - y_1$  a všechny sezónní počáteční hodnoty rovny 0 pro aditivní a 1 pro multiplikativní verzi. Tímto říkáme, že se zde žádná sezónnost nevyskytuje, což nám jisto jistě nedá dobrou shodu, ale jelikož je  $\gamma = 0.4$ , tak se bude sezónní složka postupně vyvíjet. Jelikož obě z těchto verzí mají, jak aditivní, tak i multiplikativní verzi, máme již 6 různých verzí Holtovy-Wintersovy metody.

Podívejme se na to, jak rychle se jednotlivé verze dokáží přizpůsobit dané časové řadě. Z povahy dané metody víme, že postupem času se dané verze téměř vůbec neliší bez ohledu na to, jaké počáteční hodnoty jsme zvolili, když máme stejné vyhlazovací parametry a stejný druh sezónnosti. Z tohoto hlediska se může zdát, že volba počátečních hodnot nemá téměř žádný vliv na predikci. Nicméně jelikož pozorované časové řady jsou konečné, a ty naše mají pouze 100 pozorování, tak i například predikce se může celkem lišit. Nemluvně o tom, že správná volba počátečních hodnot nám může zaručit celkem přesnou predikci i pro výrazně méně dlouhé časové řady.

Z důvodu přehlednosti se v tomto případě graficky podíváme pouze na první časovou řadu v obrázku 2.8. Ostatní vyhodnotíme pouze pomocí  $RMSE^*$ , která se vypočítá stejným vzorcem, jako klasická  $RMSE$ , avšak v tomto případě nepočítáme s predikcí, ale s vyrovnanými hodnotami, které nám metoda dala pomocí regresivních rovnic.  $RMSE^*$  tedy můžeme brát jako kvantifikátor toho, jak dobře se dokázaly jednotlivé verze aproximovat reálnou časovou řadu. Výsledky můžeme vidět v tabulce 2.6.



Obrázek 2.8: Graf první časové řady a její vyrovnané hodnoty dle jednotlivých počátečních hodnot

Metody	(1)	(2)	(3)	(4)
Teoretická aditivní	0.22	65	0.0009	0.88
Teoretická multiplikativní	0.25	58	0.0009	0.75
Sofwarová aditivní	0.28	63	0.0010	0.88
Sofwarová multiplikativní	0.29	63	0.0011	0.80
Primitivní aditivní	0.34	212	0.0027	1.14
Primitivní multiplikativní	0.35	217	0.0027	1.08

Tabulka 2.6:  $RMSE$  verzí Holtovy-Wintersovy metody na jednotlivé časové řady

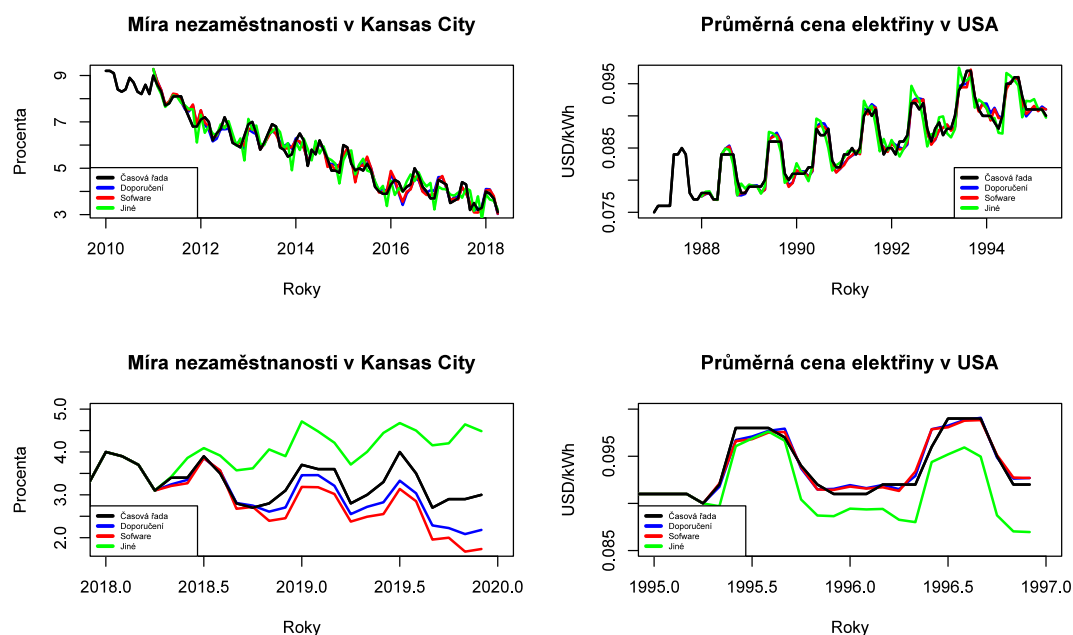
V obrázku 2.8 můžeme vidět, že jednotlivé verze pro aditivní a multiplikativní verze se příliš neliší, což nám potvrzují hodnoty  $RMSE$  pro ostatní časové řady. Pokud bychom je chtěli porovnat, tak se dozvíme, že pro časovou řadu (1) je vždy aditivní verze lepší, naopak pro časovou řadu (2) a (4) jsou právě ty multiplikativní lepší. A pro časovou řadu (3) jsou téměř stejné.

Primitivní verze se ukázaly jako nejhorší, což jsme čekali. V grafu si můžeme všimnout, že to této metodě trvalo asi do začátku roku 2015, než vidíme obdobné

sezónní chování jako u ostatních verzí. Naopak jako nejlepší se ukázaly ty teoretické, což rozhodně může působit jako překvapení, ale pořád nemůžeme usuzovat, že teoretické jsou lepší než softwarové, jelikož jejich rozdíly nejsou příliš velké a jedná se pouze o čtyři časové řady. Taktéž pokud se podíváme na graf pro první časovou řadu, tak můžeme vidět, že v počátku byly verze s námi vypočítanými počátečními hodnotami více nepřesné a příliš přeceňovali velikost sezónní složky. Avšak ověřili jsme, že počáteční hodnoty zvolené dle teoretické části jsou velmi dobré a není žádný problém je používat, pokud nemáme po ruce nějaký software, který nám je vypočítá sám.

## Vyhlazovací konstanty

Nyní podrobíme zkoušce naše doporučení o zvolení vyhlazovacích konstant. Zvolíme  $\alpha = 0.4$ ,  $\beta = 0.1$  a  $\gamma = 0.4$  a budeme je porovnávat se softwarovým algoritmem na jejich odhad. Taktéž si zde zvolíme nějakou náhodnou kombinaci, abychom si ilustrovali, že špatný odhad může mít za následek velmi špatnou shodu i predikci. Zvolme například  $\alpha = 0.8$ ,  $\beta = 0.5$  a  $\gamma = 0.1$ . Kvůli přehlednosti provedeme porovnání pouze na první a třetí časové řadě a porovnávat je budeme zároveň dle RMSE\* (obrázek 2.9 a tabulka 2.7) a taktéž dle přesnosti predikce pomocí MAE, MAPE a RMSE (obrázek 2.9 tabulka 2.8).



Obrázek 2.9: Grafy jednotlivých verzí a predikce pro různé volby vyhlazovacích konstant Holtovy-Wintersovy metody pro časové řady (1) a (3)

Zde můžeme vidět, že naše doporučení pro volbu vyhlazovacích konstant obstálo velmi dobře vůči softwaru. Dokonce jsme v přesnosti predikce software předčili. Taktéž můžeme vidět, že pro náhodně zvolené hodnoty dostaneme velice špatné výsledky, což jen svědčí o tom, že je důležité si dávat pozor na správnou volbu vyhlazovacích parametrů.

Metody	(1)	(3)
Doporučení	0.28	0.0010
Software	0.26	0.0009
Jiné	0.40	0.0015

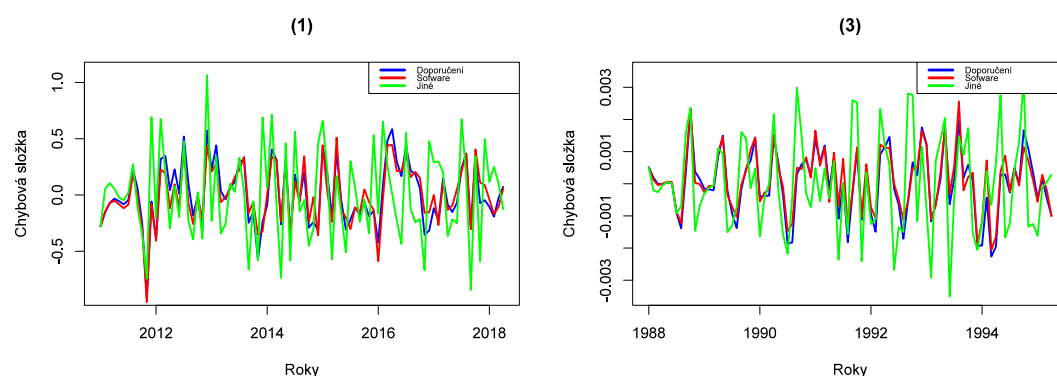
Tabulka 2.7: Tabulka RMSE\* pro volbu vyhlazovacích konstant Holtovy-Wintersovy metody

Číslo řady	Metoda	MAE	MAPE	RMSE
(1)	Doporučení	0.33	10.4 %	0.42
	Software	0.52	16.6 %	0.64
	Jiné	0.90	29.3 %	1.00
(3)	Doporučení	0.0006	0.7 %	0.0008
	Software	0.0007	0.7 %	0.0008
	Jiné	0.0029	3.1 %	0.0033

Tabulka 2.8: Tabulka přesnosti predikce pro volbu vyhlazovacích konstant Holtovy-Wintersovy metody

Pokud se podíváme na rozdíly mezi MAE a RMSE, tak můžeme vidět, že predikce pro Holtovu-Wintersovu metodu jsou velmi stabilní a nejsou zde žádné velké výkyvy v její přesnosti.

Analyzujme si ještě chybovou složku pro různé volby vyhlazovacích konstant. Jak vidíme v obrázku 2.10, tak chybová složka je téměř stejná pro naše doporučení a pro softwarovou volbu, avšak výrazně se liší od náhodné volby, která má navíc větší rozptyl, což lze vyčíst i z vyšší hodnoty RMSE\*. Pokud se však podíváme na velikost toho rozptylu, tak ten se zdá po celou dobu časové řady konstantní pro všechny volby vyhlazovacích konstant. Taktéž průměry jsou velmi blízké nule. Pokud se podíváme na první časovou řadu, lze vidět pro doporučení a softwarovou volbu náznak porušení nezávislosti, avšak oproti chybovým složkám z předchozích metod jde jistě o výrazné zlepšení.



Obrázek 2.10: Chybová složka pro jednotlivé řady dle Holtovy-Wintersovy metody pro různé hodnoty vyhlazovacích konstant

Závěr je tedy takový, že pokud bychom zde nemohli využít software, který

počáteční hodnoty a vyhlazovací parametry vypočítá sám, tak nedostaneme o nic horší výsledky, než kdybychom použili námi navrhované hodnoty.

Celkově, co se týče Holtovy-Wintersovy metody, tak musíme vyzdvihnout velkou flexibilitu, která se odráží ve velkém počtu vstupních hodnot. Ty nám pomohou vytvořit model, který přesně odpovídá zadané řadě. Výhodou je taky jeho univerzální použitelnost na velkou většinu situací a je jeden z častěji využívaných přístupů k očištění sezónnosti časové řady a taktéž následné predikci této řady. Navíc jej lze najít ve většině softwarů. Nevýhodou však může být jeho početní náročnost.

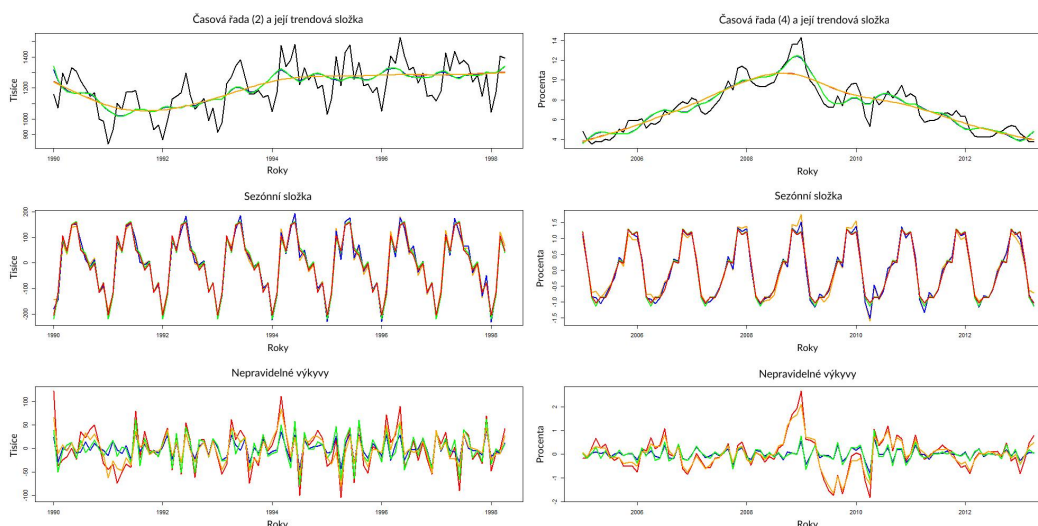
### 2.3.4 Schlichtova metoda

U Schlichtovy metody se podívejme, jak jednotlivé výběry konstant  $\alpha$ ,  $\beta$  a  $\gamma$  mění rozklad časové řady. Toto můžeme testovat například na řadě (2) a (4), které se zdají jako nejzajímavější, jelikož možná nemají tolik pravidelnou sezónnost a také nejsou tak lineární jako časové řady (1) a (3).

Budeme porovnávat 4 různé kombinace konstant dle tabulky 2.9.

Kombinace	Barevné označení	$\alpha$	$\beta$	$\gamma$
1	Červená	1000	1000	1000
2	Modrá	1	1	1
3	Zelená	1	1000	1000
4	Oranžová	1000	1	1

Tabulka 2.9: Kombinace hodnot parametrů u Schlichtovy metody



Obrázek 2.11: Schlichtova metoda použita na časové řady (2) a (4)

Můžeme vidět, že hodnoty  $\beta$  a  $\gamma$  nemají téměř žádný vliv na vzhled trendové složky, což je však něco, co jsme předpokládali. Vidíme, že pro  $\alpha = 1000$  je trend velice hladký, ignorující jakékoliv výkyvy. Pro  $\alpha = 1$  je trend taktéž celkem

hladký, ale už můžeme vidět, že na něj mají velký vliv jednotlivé jednosezónní výkyvy, což samozřejmě může být výhodou, záleží čistě na naší představě toho, co bychom rádi dostali jako výstup.

Podobně jako u trendu, tak na tvar křivky sezónnosti má hodnota parametru  $\alpha$  minimální vliv. Všimněme si, že pro  $\beta = \gamma = 1$  je tvar sezónnosti pravidelný a změny mezi jednotlivými sezónami jsou téměř nulové. Nicméně i pro  $\beta = \gamma = 1$  je sezónnost velice pravidelná, což tedy znamená, že v dané časové řadě je sezónnost doopravdy velice pravidelná. Avšak mírné změny mezi jednotlivými sezónami můžeme pozorovat. Například velikost maxima v jednotlivých sezónách je celkem proměnlivá pro obě časové řady. U poslední časové řady také můžeme vidět proměnlivé chování v okolí minima v různých sezónách.

Charakter chybové složky je pro dvě prezentované časové řady odlišný. Časová řada (4) má viditelně největší hodnoty chybové složky v okolí zlomu časové řady, tedy v období, kdy nastává změna trendu dané řady. U časové řady (2) žádný takový jednorázový nárůst chybové složky nepozorujeme. Můžeme však pozorovat, že případ, kdy jsme zvolili všechny parametry rovny 1000, má nejvyšší hodnoty chybové složky. Nicméně i kdybychom zvolili  $\alpha = 1000$ , ale další dvě konstanty pouze 1, tak co se týče velikosti hodnot chybové složky, tak jsme si velice nepolepšili. Toto jen potvrzuje naše tvrzení, že obě časové řady mají celkem pravidelnou sezónnost, protože obě kombinace, které chtějí vysokou pravidelnost sezónní složky, mají velice malé hodnoty chybové složky. Pokud by byly sezónní výkyvy reálně více nepravidelné, ale my bychom chtěli rozklad na velice pravidelnou sezónní složku, jistě bychom vysoké hodnoty chybové složky.

Průměr chybové složky je u všech případů velice blízký nule. Její rozptyl se zdá konstantní u časové řady (4) a taktéž u časové řady (2) pro volbu  $\alpha = 1$ . Platí, že pro vyšší hodnoty parametrů je rozptyl vyšší a taktéž nezávislost se zdá být více porušena, což je přesně to, čeho se vzdáváme, požadujeme-li, aby trendová složka byla velmi vyhlazená a sezónní složka velice pravidelná.

Celkově se tedy Schlichtova metoda ukazuje jako užitečný nástroj, který má celkem jednoduché ovládání, co se týče zvolení vstupních hodnot parametrů. Jednoduše, pokud chceme velmi hladký trend, budeme volit vysoké  $\alpha$ , pokud chceme pravidelnou a neměnnou sezónnost budeme naopak volit vysoké  $\beta$  a  $\gamma$ , což přesně potvrzuje to, co jsme již popisovali v teoretické části, ve které jsme také rozebrali možnost odhadu těchto parametrů. Avšak za vyšší vyhlazenost trendové složky a vyšší pravidelnost té sezónní musíme zaplatit větší chybovou složkou. Nevýhoda Schlichtovy metody je rozhodně neexistující teorie pro predikci pomocí této metody.

### 2.3.5 ETS Metody

V této sekci se podíváme na to, jak software aplikuje ETS metody na naše časové řady pomocí funkce `ets()`. Ta nám dokáže vykreslit grafy jednotlivých složek, které tvoří časovou řadu (obrázek 2.12). Taktéž nám podá informaci o všech parametrech, které při opakovaném použití regresních rovnic do procesu vstupují. Jelikož tato funkce dokáže vykreslovat i predikční intervaly a další, tak kromě metody vybírá i přesný model dle charakteru chybové složky, což však nemá vliv na výpočet predikce. Další důležité hodnoty, které dostáváme aplikací funkce `ets()`, zapíšeme ve formě tabulky (2.10).

Časová řada	Model	$\alpha$	$\beta$	$\gamma$	RMSE*
(1)	(A,A,A)	0.599	0	0	0.20
(2)	(A,N,A)	0.343	-	0	54
(3)	(M,A,M)	0.340	0	0	0.0010
(4)	(M,A,M)	0.539	0.02	0	0.08

Tabulka 2.10: Jednotlivé aplikace ETS metod na časové řady

Letným pohledem na RMSE\*, které jsme vypočítali u jednotlivých metod, můžeme vidět, že máme velice dobrou shodu pro všechny časové řady. Co dále můžeme vidět v tabulce je, že hodnoty  $\gamma$  jsou rovny 0 pro všechny časové řady, což znamená, že sezónnost je plně pravidelná a neměnná, což sice potvrzuje, že všechny časové řady silně vykazují sezónnost, avšak může následně vytvářet nepřesné odhady, pokud by se sezónnost měnila jen minimálně. Taktéž hodnoty  $\beta$  jsou ve dvou případech nulové a v jednom velice malé, což nám tedy říká, že trend je velice lineární.

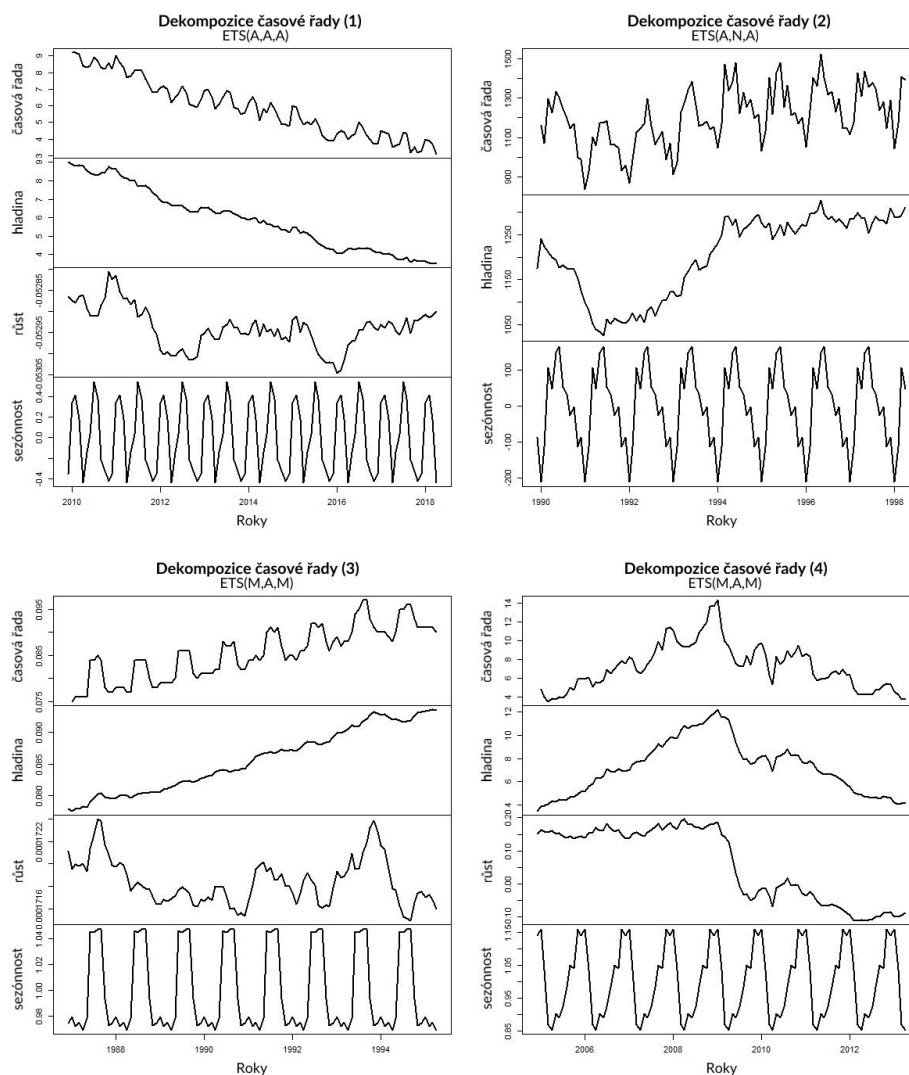
Pojďme se nyní postupně podívat na jednotlivé řady. U první časové řady vidíme, že ji naše metoda považuje za aditivní, tedy splňuje definici 3. Ve své podstatě zde vlastně používáme aditivní Holtovu-Wintersovu metodu. Vidíme, že hladina postupně klesá velmi lineárně, avšak ke konci je patrná mírná změna trendu. Pro růstovou složku můžeme vidět, že se po celou dobu drží mezi hodnotou  $-0.0527$  a  $-0.053$  a tedy je téměř konstantní, i když to z obrázku není na první pohled úplně patrné. Můžeme vidět, že sezónnost je, jak jsme si již řekli, velice pravidelná. Taktéž si lze povšimnout, že změna sezónního výkyvu zde nastává dvakrát za sezónu. Toto chování nevidíme u žádné jiné časové řady.

Pokud se podíváme na druhou časovou řadu, tak nás překvapí, že nejlepšího výsledku, co se týče informačního kritéria, dosahuje metoda, která nemá růstovou složku. Pokud se podíváme na hladinu, tak vidíme, že v první části se velmi mění, avšak v druhé půlce je již téměř konstantní. Toto může být potvrzení toho, že tato časová řada doopravdy nemá růstovou složku. U této metody jsme taktéž odhadli celkem malou hodnotu  $\alpha$ , což znamená, že hladina má být spíše konstantní, než že by se měla přizpůsobovat různým výkyvům.

Vidíme, že u časové řady (3) dostáváme celkem malé  $\alpha$  a nulové  $\beta$ . To znamená, že hladina svůj směr mění pomalu a nepřizpůsobuje se různým výkyvům. To lze taktéž vidět z toho, že růstová složka je téměř konstantní. Ze sezónnosti se dozvídáme, že při změně směru sezónních výkyvů nastává strmý pád či strmý růst asi o 6 %.

Poslední časová řada je stejně jako ta předchozí multiplikativní dle definice 3. Můžeme i říci, že vlastně používáme multiplikativní Holtovu-Wintersovu metodu. Pro tuto časovou řadu je  $\beta$  rovno 0.02, a tedy je zde možná pozvolná změna růstové složky. Toto můžeme jednoznačně vidět, když se v roce 2009 růstová složka propadne z  $+0.15$  na  $-0.05$  a až dokonce k  $-0.1$ . Následkem je nelineárnost složky hladiny, což nám tedy potvrzuje, že trend časové řady není lineární.

Pokud se podíváme na chybovou složku (obrázek 2.13), tak můžeme vidět, že velice dobře splňuje všechny předpoklady. Průměry jsou relativně blízké nule, rozptyly jsou konstantní po celou délku časové řady a i nezávislost je porušena jen minimálně. Jemné porušení můžeme sledovat například na počátku druhé časové řady. Avšak celkově musíme být s charakterem chybové složky u ETS metod velice

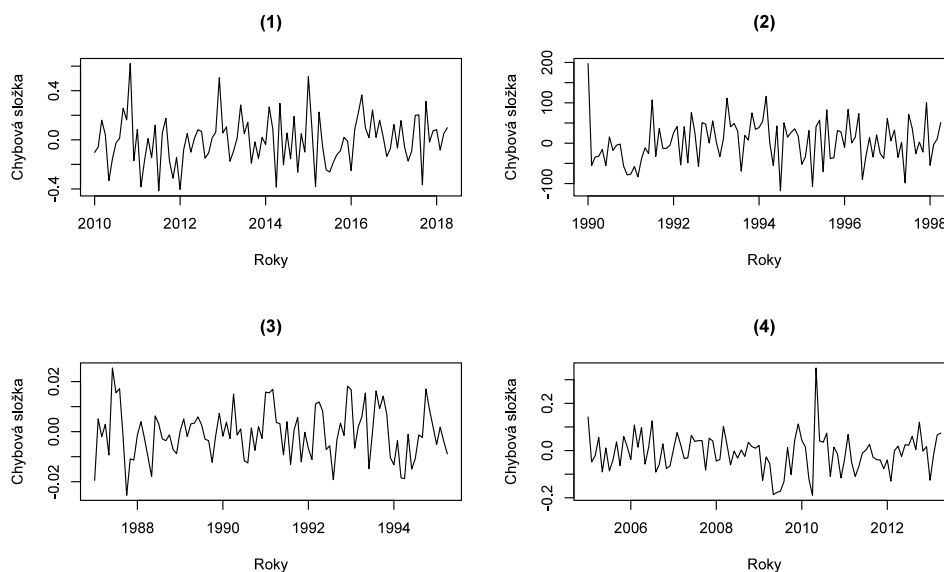


Obrázek 2.12: Dekompozice jednotlivých časových řad pomocí ETS metod

spokojeni.

Největší předností ETS metod je právě to velké množství metod, ze kterých si můžeme vybrat. Když k tomu přidáme ještě možnost volby vyhlazovacích parametrů a počátečních hodnot, tak dostáváme nejflexibilnější metodu. Protože tyto hodnoty odhadujeme pomocí maximální věrohodnosti a na výběr nejlepší metody používáme informační kritérium, je práce s touto metodou s pomocí softwaru velice jednoduchá. Další výhodnou je velmi uspokojivý charakter chybové složky. Nevýhodou je jednoznačně nemožnost počítat s těmito metodami bez přístupu ke správnému softwaru.





Obrázek 2.13: Chybová složka pro jednotlivé řady dle ETS metod

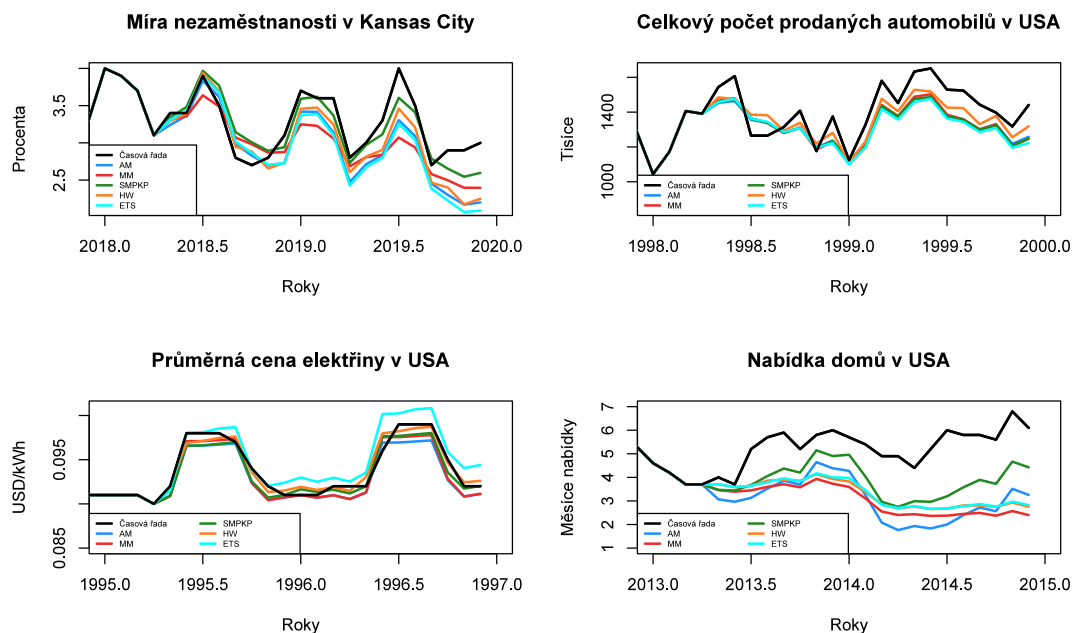
## 2.4 Komparace jednotlivých přístupů

V této sekci se budeme zabývat porovnáním jednotlivých přístupů z hlediska přesnosti predikce. Využijeme přitom znalosti, které jsme zjistili v předchozích kapitolách. Samozřejmě nemůžeme využít znalost toho, jak si jednotlivé verze metod vedly v predikci, které jsme již prováděli. Avšak řekli jsme si a ukázali na grafech (obrázek 2.6), že pro modelování sezónnosti pomocí kvalitativní proměnné (zkratka SMPKP) je dosažení  $f(t) = 1 - 0.99^t$  dobrá shoda pro první tři časové řady. Pro poslední jsme použili po částech lineární funkci, která měla uzly dle tabulky 2.4. Pro Holtovu-Wintersovu metodu využijeme verzi, která měla pro jednotlivé časové řady nejmenší RMSE\* dle tabulky 2.6. Můžeme však vidět, že ve všech případech je nejlepší verze s námi vypočítanými počátečními hodnotami. Vyhlazovací konstanty použijeme dle doporučení, a to  $\alpha = 0.4$ ,  $\beta = 0.1$  a  $\gamma = 0.4$ . Pro jednoduché přístupy k sezónnosti budeme porovnávat obě verze, jak aditivní, tak i multiplikativní (zkratka AM a MM). A ETS metody budeme aplikovat pomocí softwarové funkce.

Nyní se tedy pojdme podívat na to, jak si jednotlivé přístupy vedly v obrázku 2.14 a tabulce 2.11.

Pokud se podíváme na jednoduché přístupy k sezónnosti, tak vidíme, že rozdíly mezi aditivní a multiplikativní metodou jsou velice malé. Co se týče jejich porovnání s ostatními metodami, tak vidíme, že většinou patří k těm horším predikcím, ale jsou často celkem blízko ostatních metod. Pro první časovou řadu se model multiplikativní metody nehodí, což jsme si již dříve řekli.

Přístup modelování sezónnosti pomocí kvalitativní proměnné zde slaví velký úspěch. Pro první a poslední časovou řadu má nejlepší predikce. Dobrou predikci má i pro třetí časovou řadu a pro druhou jsou výsledky celkem uspokojivé. Ukazuje se tedy, že pro první metodu trend rozhodně není lineární, ale postupem času rychlost poklesu zpomaluje. U čtvrté časové řady jsme benefitovali z toho, že trendovou složku pro následující období jsme odhadovali v podstatě z posledních



Obrázek 2.14: Grafy predikcí dle jednotlivých přístupů pro jednotlivé časové řady

pár pozorování, a tedy jsme dokázali alespoň trochu zachytit výraznou změnu trendu. Nevýhodou je zde znovu to, že jsme my museli vybírat tvar trendové křivky a od toho se odvíjel úspěch predikce. Kdybychom špatně zvolili trend, například bychom si mysleli, že nejlepší by byla predikce s lineárním trendem, tak bychom naopak dostali velmi neuspokojivé výsledky.

Pokud se díváme pouze na tyto čtyři časové řady, tak se Holtova-Wintersova musí zdát jako ta nejlepší. Pro dvě časové řady je její predikce ta nejpřesnější a pro další dvě řady je taktéž celkem přesná. Vždy je přesnější než jednoduché metody. Výhodou je, že není žádný náš vliv a funkce dělá vše sama za sebe podle předem naprogramovaných příkazů.

Metody ETS se zde zdají jako veliké zklamání. Ve třech případech jsme dostali nejhorší predikci. Například pro časovou řadu (3) je její predikce výrazně vzdálena všem ostatním. Zvláštní je to také, protože pro tři ze čtyř časových řad vlastně používáme Holtovu-Wintersovu metodu, kde jsme dané parametry odhadli pomocí metody maximální věrohodnosti. Avšak ve dvou ze tří těchto případů je naše Holtova-Wintersova metoda výrazně lepší.

Pokud se podíváme na konzistenci našich predikcí, tak vidíme, že všechny metody mají zhruba stejnou stabilitu. Rozdíly jsou spíše mezi jednotlivými časovými řadami, kdy pro druhou a čtvrtou časovou řadu jsou predikce více konzistentní a pro první a třetí méně.

Z našich výsledků tedy můžeme usoudit, že Holtova-Wintersova metoda je velice dobrým nástrojem pro velkou škálu podobných časových řad, jako se kterými jsme pracovali.

Číslo řady	Metoda	MAE	MAPE	RMSE
(1)	Aditivní	0.33	10.5 %	0.40
	Multiplikativní	0.33	9.9 %	0.40
	SMPKP	0.18	5.7 %	0.22
	Holt-Winters	0.29	9.1 %	0.35
	ETS	0.37	11.6 %	0.45
(2)	Aditivní	108	7.4 %	119
	Multiplikativní	107	7.4 %	118
	SMPKP	109	7.5 %	121
	Holt-Winters	84	5.8 %	92
	ETS	119	8.2 %	131
(3)	Aditivní	0.0012	1.2 %	0.0013
	Multiplikativní	0.0010	1.1 %	0.0012
	SMPKP	0.0008	0.9 %	0.0010
	Holt-Winters	0.0006	0.7 %	0.0008
	ETS	0.0013	1.3 %	0.0016
(4)	Aditivní	2.36	43.3 %	2.53
	Multiplikativní	2.41	43.2 %	2.59
	SMPKP	1.53	28.0 %	1.66
	Holt-Winters	2.13	38.0 %	2.32
	ETS	2.11	37.6 %	2.30

Tabulka 2.11: MAE, MAPE a RMSE jednotlivých přístupů pro jednotlivé časové řady

# Závěr

V této práci jsme shrnuli některé velice efektivní přístupy k sezónnímu očišťování ekonomických časových řad a jejich následnou predikcí. Přesně jsme popsali jednotlivé kroky k očištění u jednoduchých přístupů, taktéž jsme si uvedli model využívající kvalitativní proměnnou. Dále jsme si představili velice používanou Holtovu-Wintersovu metodu, jak pro aditivní, tak i pro multiplikativní dekompozici. Poté jsme přistoupili k očištění sezónnosti z hlediska optimalizace a popsali jsme si Schlichtovu metodu. A jako poslední v teoretické části jsme si uvedli robustní teorii k ETS metodám.

V praktické části jsme navázali na teorii a ukázali jsme použití představených metod v praxi na vybrané ekonomické časové řady. Upozornili jsme na přednosti a nedostatky u jednotlivých metod, které byly podpořeny výpočty i grafy. Také jsme se podívali na vlastnosti chybové složky. Následně jsme přešli ke komparaci jednotlivých časových řad pomocí přesnosti predikce.

Hlavní závěrem této praktické části je, že Holtova-Wintersova metoda dosahovala nejlepších výsledků s benefitem toho, že uživatel zde nemusí, ale může do procesu očišťování a predikování zasahovat.

# Seznam použité literatury

- CIPRA, T. (2013). *Finanční Ekonometrie*. Ekopress, Praha. ISBN 978-80-86929-93-4.
- HYNDMAN, R. J. a KHANDAKAR, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, **27**(3). doi: 10.18637/jss.v027.i03.
- HYNDMAN, R. J., KOEHLER, A., ORD, K. a SNYDER, R. (2008). *Forecasting with exponential smoothing : the state space approach*. Springer, Berlin. ISBN 978-3-540-71916-8.
- RENCHER, A. C. a SCHAALJE, G. B. (2000). *Linear models in statistics*. John Wiley & Sons, New York. ISBN 0-471-31564-8.
- SCHLICHT, E. (1981). A seasonal adjustment principle and a seasonal adjustment method derived from this principle. *Journal of the American Statistical Association*, **76**(374), 374–378. doi: 10.1080/01621459.1981.10477656.
- SVETUNKOV, I. (2022). Forecasting and analytics with adam. URL <https://openforecast.org/adam/>.
- WATSON, N. (2012). Using mean absolute error for forecast accuracy. URL <https://canworksmart.com/using-mean-absolute-error-forecast-accuracy/>.

# Seznam obrázků

2.1	Grafy jednotlivých časových řad . . . . .	20
2.2	Časové řady očištěné od sezónní složky dle jednoduchých přístupů	23
2.3	Grafy predikcí dle jednoduchých přístupů a reality pro jednotlivé časové řady . . . . .	24
2.4	Chybová složka pro jednotlivé řady dle přístupu jednoduché aditivní a multiplikativní dekompozice . . . . .	25
2.5	Grafy predikcí dle metody SMPKP a reality pro jednotlivé časové řady . . . . .	26
2.6	Grafy trendové složky dle jednotlivých SMPKP přístupů pro jednotlivé časové řady . . . . .	27
2.7	Chybová složka pro jednotlivé řady dle modelu SMPKP . . . . .	28
2.8	Graf první časové řady a její vyrovnané hodnoty dle jednotlivých počátečních hodnot . . . . .	30
2.9	Grafy jednotlivých verzí a predikce pro různé volby vyhlazovacích konstant Holtovy-Wintersovy metody pro časové řady (1) a (3) . .	31
2.10	Chybová složka pro jednotlivé řady dle Holtovy-Wintersovy metody pro různé hodnoty vyhlazovacích konstant . . . . .	32
2.11	Schlichtova metoda použita na časové řady (2) a (4) . . . . .	33
2.12	Dekompozice jednotlivých časových řad pomocí ETS metod . . . .	36
2.13	Chybová složka pro jednotlivé řady dle ETS metod . . . . .	37
2.14	Grafy predikcí dle jednotlivých přístupů pro jednotlivé časové řady	38

# Seznam tabulek

1.1	Umělé proměnné reprezentující sezónnost . . . . .	8
1.2	Možné kombinace povah růstové a sezónní složky . . . . .	14
1.3	Počet parametrů pro jednotlivé ETS metody . . . . .	15
1.4	Popis stavového prostoru ETS modelů s aditivní chybovou složkou	17
1.5	Popis stavového prostoru ETS modelů s multiplikatívní chybovou složkou . . . . .	18
1.6	Vzorce pro rekurzivní kalkulace a predikce ETS modelů . . . . .	19
2.1	Popis jednotlivých časových řad . . . . .	21
2.2	Základní popisné charakteristiky jednotlivých časových řad . . . .	21
2.3	MAE, MAPE a RMSE jednoduchých přístupů pro jednotlivé ča- sové řady . . . . .	22
2.4	Uzly pro jednotlivé časové řady . . . . .	26
2.5	MAE, MAPE a RMSE přístupů SMPKP pro jednotlivé časové řady	27
2.6	RMSE verzí Holtovy-Wintersovy metody na jednotlivé časové řady	30
2.7	Tabulka RMSE* pro volbu vyhlazovacích konstant Holtovy-Wintersovy metody . . . . .	32
2.8	Tabulka přesnosti predikce pro volbu vyhlazovacích konstant Holtovy- Wintersovy metody . . . . .	32
2.9	Kombinace hodnot parametrů u Schlichtovy metody . . . . .	33
2.10	Jednotlivé aplikace ETS metod na časové řady . . . . .	35
2.11	MAE, MAPE a RMSE jednotlivých přístupů pro jednotlivé časové řady . . . . .	39