



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA STROJNÍHO INŽENÝRSTVÍ
FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY
INSTITUTE OF MATHEMATICS

ANALÝZA A PREDIKCE ČASOVÝCH ŘAD VE FINANČNICTVÍ

ANALYSIS AND FORECASTING
OF TIMESERIES IN FINANCING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

ERIK MAŇÁSEK

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. PAVEL HRABEC, Ph.D.

BRNO 2024

Zadání bakalářské práce

Ústav: Ústav matematiky
Student: **Erik Maňásek**
Studijní program: Matematické inženýrství
Studijní obor: bez specializace
Vedoucí práce: **Ing. Pavel Hrabec, Ph.D.**
Akademický rok: 2023/24

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Analýza a predikce časových řad ve finančnictví

Stručná charakteristika problematiky úkolu:

Úkolem autora bude nastudovat základní nástroje modelování a predikování časových řad. A ty potom aplikovat na vybrané časové řady z oblasti finančnictví.

Cíle bakalářské práce:

- seznámení se s matematickým aparátem vhodným pro modelování časových řad ve finančnictví
- aplikace těchto nástrojů na vybraná data z oblasti finančnictví
- diskuse výsledků

Seznam doporučené literatury:

MONTGOMERY, D. C. Forecasting and Time Series Analysis (2nd ed). McGraw-Hill, 1990.

LAI, T. L. & XING, H. Statistical models and methods for financial markets. Springer, 2008.

Abstrakt

Tématem této bakalářské práce je technická analýza finančních časových řad akcií na burze. Práce rozebírá metodou dekompozice na jednotlivé složky, způsoby analýzy a predikce těchto složek. Dále zde nalezneme porovnání použitých metod, jejich zhodnocení a nástin některých jejich vlastností, jak těch dobrých, tak těch špatných.

Abstract

The topic of this bachelor thesis is the technical analysis of stock market financial time series. The thesis discusses the method of decomposition into individual components, methods of analysis and prediction of these components. It also provides a comparison of the methods used, their evaluation and an outline of some of their characteristics, both good and bad.

Klíčová slova

lineární regrese, exponenciální vyrovnaní, EMA, spektrální analýza, finanční časové řady, akcie

Keywords

linear regression, exponential moving average, EMA, spectral analysis, financial time series, stocks

Prohlašuji, že jsem bakalářskou práci *Analýza a predikce časových řad ve finančnictví* vypracoval samostatně pod vedením Ing. Pavla Hrabce, Ph.D. s použitím materiálů uvedených v seznamu literatury.

Erik Maňásek

Tímto bych chtěl poděkovat svému vedoucímu, Ing. Pavlu Hrabci, Ph.D, za ochotu, cenné rady a značné množství času, které mi věnoval při vedení této práce.

Erik Maňásek

Obsah

1 Úvod	12
2 Základní pojmy	13
2.1 Druhy grafů	13
2.1.1 Liniový graf	13
2.1.2 Spojnicový graf	14
2.1.3 Svíčkový graf	15
2.2 Statistické pojmy	16
2.2.1 Bílý šum	16
2.2.2 Homoskedasticita	16
2.2.3 Heteroskedasticita	16
3 Typy analýzy	17
3.1 Dekompozice časových řad	17
3.1.1 Aditivní dekompozice	20
3.1.2 Multiplikativní dekompozice	20
3.2 Boxova-Jenkinsova metodologie	21
3.3 Lineární dynamické modely	21
3.4 Spektrální analýza	21
4 Určení trendu	22
4.1 Lineární regrese	22
4.1.1 Lineární regresní model plné hodnosti	22
4.1.2 Metoda nejmenších čtverců	22
4.1.3 Prostý lineární regresní model	23
4.1.4 Obecnější lineární regresní model	24
4.1.5 Informativní testy pro volbu trendové křivky	27
4.2 Adaptivní metody	28
4.2.1 Jednoduchý klouzavý průměr	28
4.2.2 Exponenciální vyrovnání	30
5 Určení cyklické složky	35
5.1 Spektrální analýza	35
5.1.1 Cyklické chování a periodičita	35
5.1.2 Spektrální hustota	37
5.1.3 Periodogram a diskrétní Fourierova transformace	39
6 Praktická část	42
6.1 Implementace	42
6.2 Akcie Rolls Royce	42
6.2.1 Výchozí data	42
6.2.2 Výpočet trendu	43
6.2.3 Výpočet periodicity	45
6.2.4 Kompletní model	50
6.3 Akcie Admiral Group	53
6.3.1 Výchozí data	53
6.3.2 Výpočet trendu	54

6.3.3	Výpočet periodicity	55
6.3.4	Kompletní model	57
6.4	Akcie BAE systems	60
6.4.1	Výchozí data	60
6.4.2	Výpočet trendu	61
6.4.3	Výpočet periodicity	62
6.4.4	Kompletní model	64
7	Závěr	66
8	Přílohy	67

1 Úvod

Časová řada je soubor chronologicky uspořádaných pozorování. V této práci se zabývám analýzou časových řad ve finančnictví, konkrétně vývojem cen finančních instrumentů. Tyto časové řady se vyznačují tím, že jsou diskrétní a po částech ekvidistantní. Tato vlastnost je dána otevírací dobou burzy, v tomto případě se jedná o akciovou burzu London Stock Exchange (LSE) a data poskytnutá platformou Metatrader 5, která poskytuje data od pondělí do pátku v čase 10:00 - 18:30 SELČ (středoevropský letní čas), tzn. +2h GMT (Greenwich mean time).

Smyslem analýzy časových řad je především snaha vysvětlit chování v minulosti a pomocí těchto znalostí poté predikovat vývoj do budoucnosti. Předpovídání časových řad lze rozdělit dle délky předpovědi na krátkodobé v řádu několika časových period, obvykle dní až týdnů, střednědobé v řádu měsíců až dvou let a dlouhodobé na několik let dopředu. Zejména pro krátkodobé předpovědi je typické využívání technické analýzy dat, protože historická data obvykle vykazují setrvačnost bez dramatických změn za krátký časový úsek. Naproti tomu pro dlouhodobé předpovědi je důležitá fundamentální analýza, která se zabývá příčinami změn a ne pouze studiem grafických znázornění sledované časové řady [4].

Mezi základní způsoby analýzy a modelování finančních časových řad patří dekompozice časových řad, Boxova-Jenkinsova metodologie, lineární dynamické modely a spektrální analýza. V této práci se budu zabývat lineární regresi a adaptivní metodou k určení trendové složky, konkrétně exponenciálním vyrovnáním a spektrální analýzou pro určení cyklických složek stacionární beztrendové časové řady.

Práce je rozdělená do pěti částí. První částí jsou základní pojmy, kde vysvětluji některé specifické pojmy pro statistiku a finanční analýzu, které nemusejí být pro nezasvěceného čtenáře známé. Další kapitolou jsou pak typy analýzy, ve kterých se snažím shrnout základní způsoby technické analýzy časových řad. Poté přichází oddíl určení trendu, kde se blíže zabývám mnou použitými způsoby určení trendové složky. Poslední teoretickou kapitolou je určení cyklické složky, kde se zabývám základy spektrální analýzy. A nakonec je praktická část, ve které se snažím popsat své výsledky při vlastním programování daných metod a užití této technické analýzy na vybrané časové řady a nakonec formulovat závěr k tomuto tématu.

2 Základní pojmy

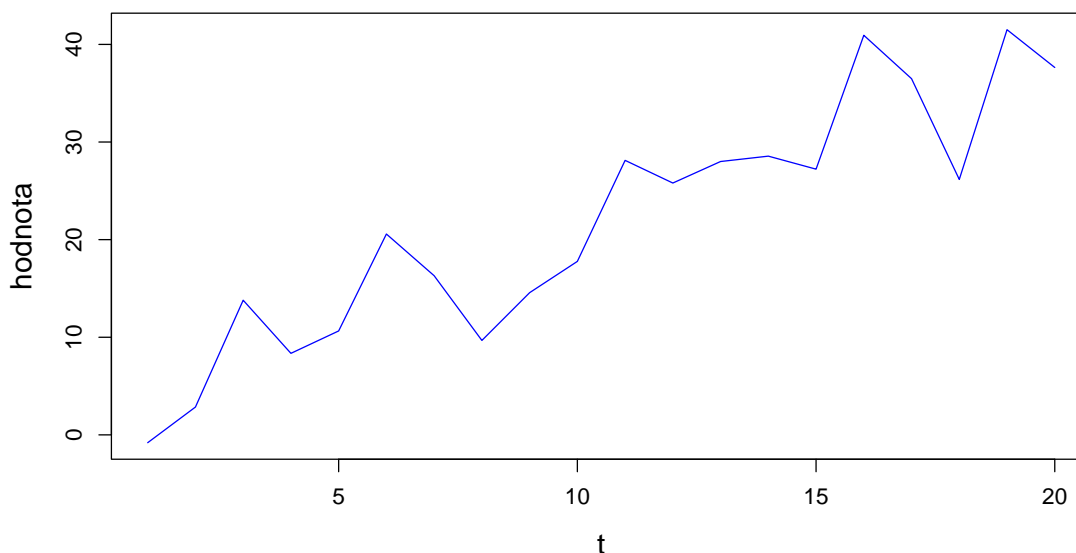
2.1 Druhy grafů

Ve finančnictví se můžeme setkat s velkým množstvím druhů grafů časových řad. Mezi ně patří zejména obyčejné liniové grafy, spojnicové grafy, svíčkové grafy a objemové histogramy.

Já jsem při své práci používal první tři zmíněné druhy, proto je zde okomentuji.

2.1.1 Liniový graf

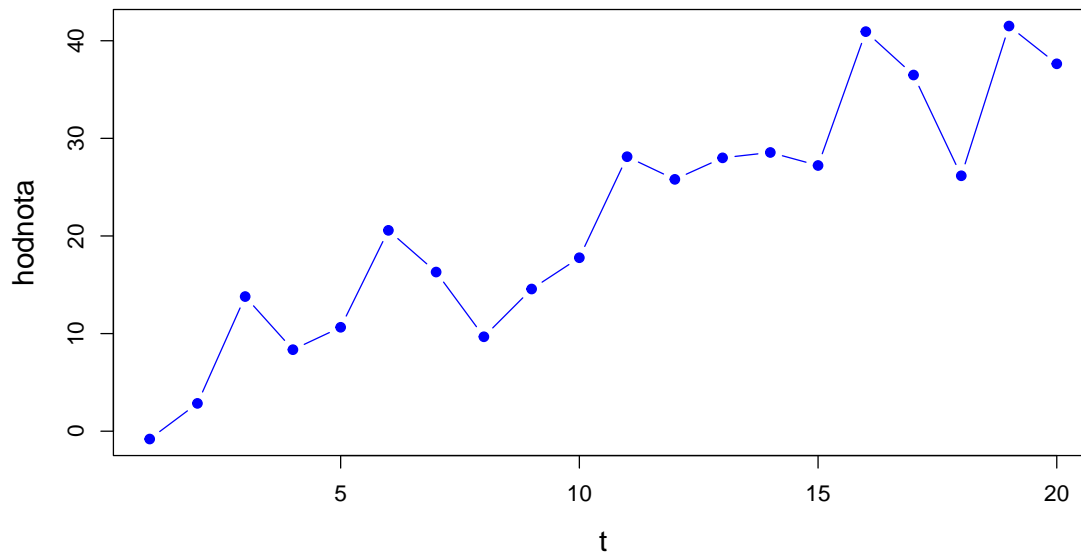
Liniový graf je základním typem grafu, který spojuje datové body pomocí úseček bez vyznačení konkrétních bodů. Tento typ grafu dokáže zobrazit pouze jeden druh ceny pro každou linii. Je platný zejména při zobrazování trendu, střední hodnoty, intervalových odhadů apod.



Obrázek 2.1: Liniový graf

2.1.2 Spojnicový graf

Spojnicový graf je dalším základním typem grafu, ale naruší od liniového grafu zobrazuje konkrétní datové body, což se může hodit zejména při analýze řídkých dat. Já jsem tento typ grafu používal pro zobrazování periodogramů při spektrální analýze.



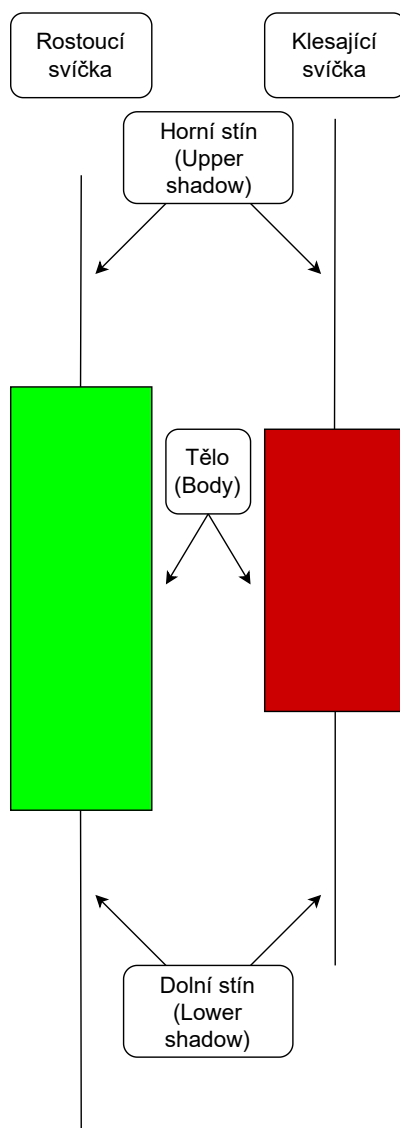
Obrázek 2.2: Spojnicový graf

2.1.3 Svíčkový graf

Posledním mnou použitým grafem je svíčkový graf. Tento typ grafu je typický pro zobrazování dat na burze, protože umožňuje zobrazovat pro každý časový interval více informací. Těmito informacemi jsou :

- **Close cena** (uzavírací cena) udává při jaké ceně se v daném časovém intervalu zavíraly ceny akcií, tzn. na jaké hodnotě cena za akcii v tomto intervalu končila.
- **Open cena** (otevírací cena) udává při jaké ceně se ceny otvíraly, tzn. počáteční hodnota v daném časovém intervalu.
- **Upper shadow** (horní stín) udává nejvyšší cenu akcie během časového intervalu.
- **Lower shadow** (dolní stín) udává nejnižší cenu akcie během časového intervalu.

Jak vidíme, svíčkové grafy poskytují lepší představu o vývoji ceny, než obyčejné liniové a spojnicové grafy, zejména při studiu grafů o vyšších frekvencích dat (denní, hodinové, minutové a další grafy).

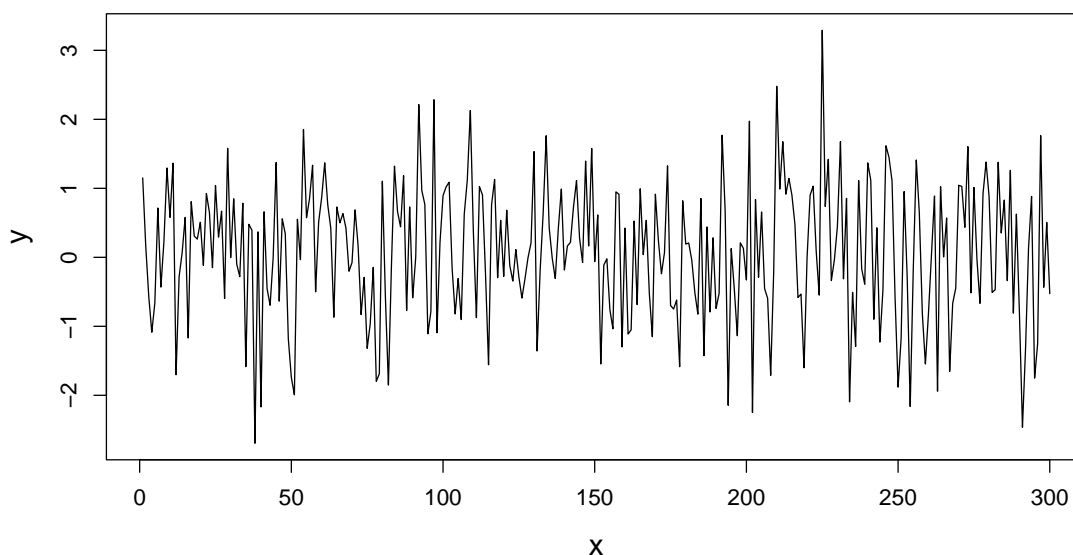


Obrázek 2.3: Svíčkový graf

2.2 Statistické pojmy

2.2.1 Bílý šum

Bílý šum je posloupnost sériově nekorelovaných náhodných proměnných s nulovou střední hodnotou a konečným rozptylem [15].



Obrázek 2.4: Bílý šum

2.2.2 Homoskedasticita

Homoskedasticita je statistický pojem, který říká, že rozptyl reziduálních hodnot závislé proměnné nezávisí na hodnotách nezávislých proměnných, konkrétně v našem případě to znamená, že se rozptyl nemění s časem [1].

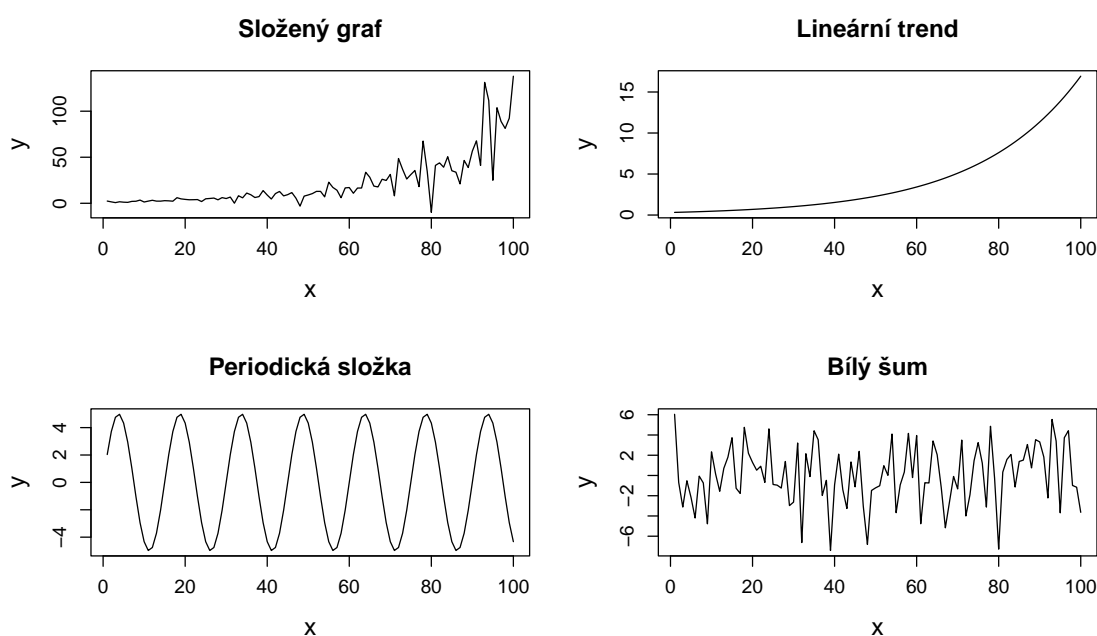
2.2.3 Heteroskedasticita

Heteroskedasticita je opakem homoskedasticity a znamená, že rozptyl reziduálních hodnot závislé proměnné na hodnotách závislých proměnných závisí, v našem případě to znamená, že se rozptyl středních hodnot mění s časem [1].

3 Typy analýzy

3.1 Dekompozice časových řad

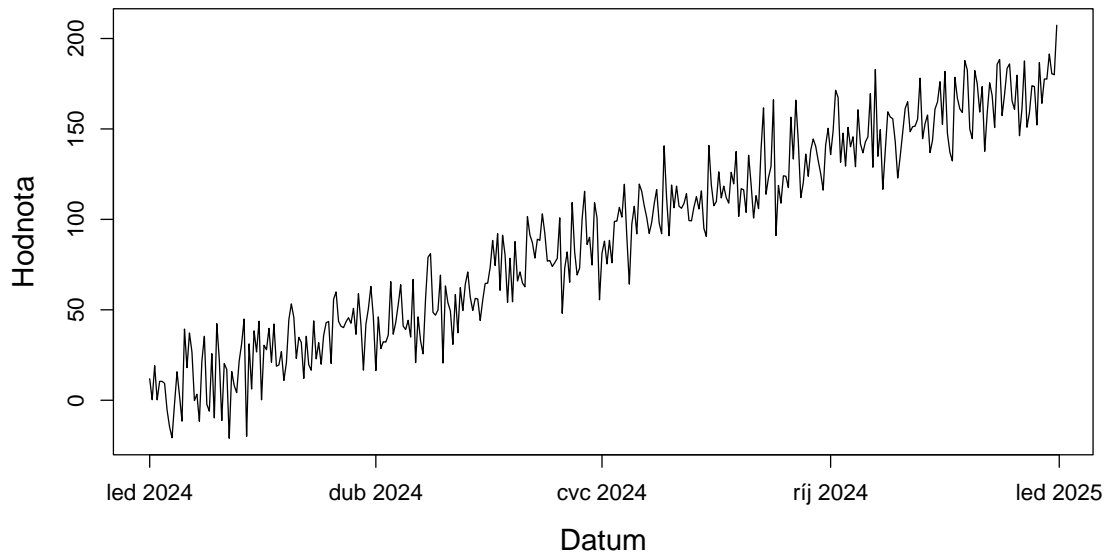
Dekompozice časové řady je rozložení časové řady na její základní složky, tj. trendovou Tr_t , periodické složky: sezónní Sz_t a cyklickou C_t a reziduální složku e_t . Důvodem k rozdělení periodických složek na sezónní a cyklickou je fakt, že délku sezónní složky umíme vysvětlit, jedná se například o výkyvy dle ročních období, měsíců atd.[4].



Obrázek 3.1: Dekompozice

Trendová složka

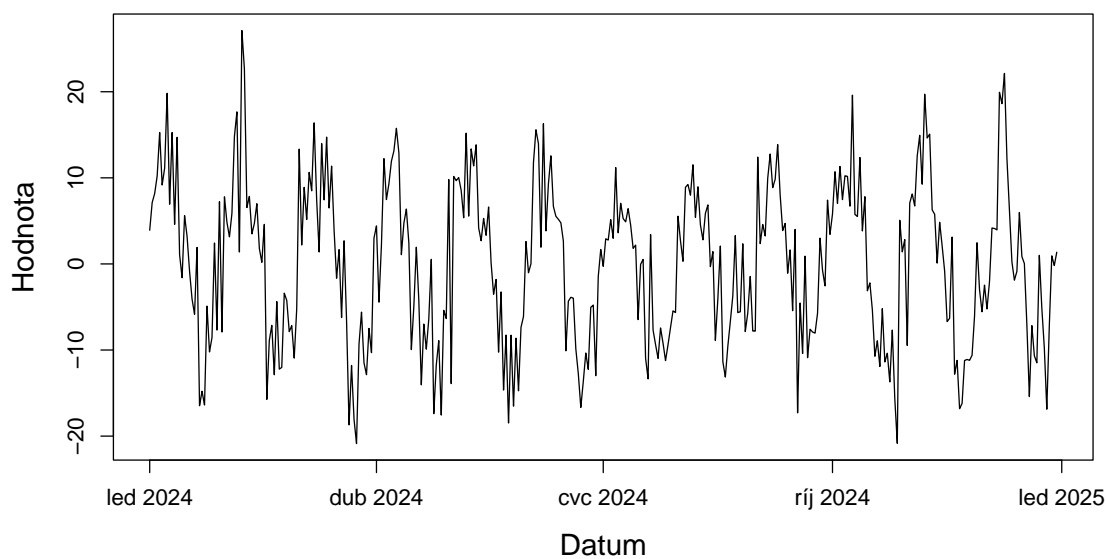
Trend ukazuje dlouhodobé změny v průměrných hodnotách časové řady, například dlouhodobý růst, nebo pokles. Při sledování trendu závisí na optice, s jakou se na časovou řadu díváme. Můžeme hledat trend například za poslední měsíc, ale i za posledních několik let. Je zřejmé, že v optice několika let budou jednotlivé měsíční trendy často kolísat a měnit svůj charakter. Tyto výkyvy v různých měsících mohou pak skončit v optice dlouhodobějšího vývoje jako periodické složky, zvláště jestliže určíme trend pomocí lineární regrese [4].



Obrázek 3.2: Časová řada s výraznou trendovou složkou

Sezónní složka

Sezónní složka popisuje periodické změny v časové řadě odehrávající se během jednoho roku, a které se každý rok opakují. Může se jednat o výkyvy způsobené změnami ročních období apod. Mají pevnou neměnnou periodu [4].



Obrázek 3.3: Časová řada s výraznou sezónní složkou

Cyklická složka

Cyklická složka naruší od sezónní nemusí mít pevnou periodu, ani neměnný charakter, může se jednat například o změny způsobené cyklickými změnami klimatu apod, kde neznáme pevnou periodu, ale pouze její průměrnou délku. Tato složka se obvykle obtížně odhaluje [4].

Reziduální složka

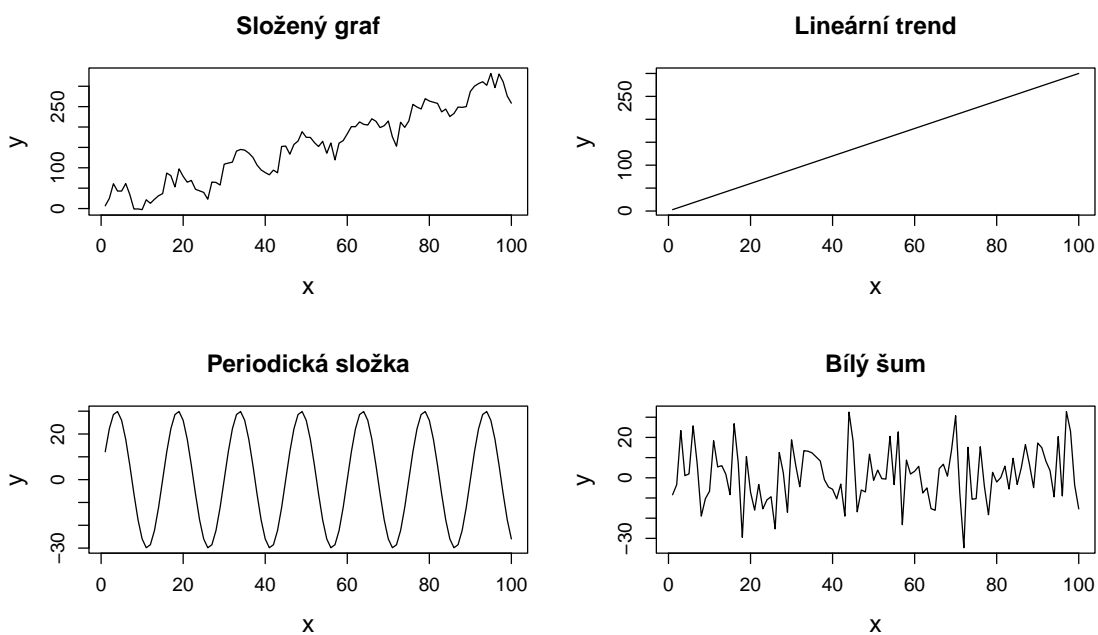
Po odstranění trendové a periodických složek zbývá složka reziduální. Jedná se o nesystematické fluktuace. Tato složka pokrývá chyby v měření, při zaokrouhlení atd. Snahou při dekompozici je, aby reziduální složka byla tzv. bílým šumem [4].

3.1.1 Aditivní dekompozice

Při aditivní dekompozici předpokládáme, že řada Y_t lze rozložit ve tvaru

$$Y_t = Tr_t + Sz_t + C_t + e_t. \quad (3.1)$$

Je zřejmé, že v tomto tvaru mají všechny složky stejnou jednotku [4].



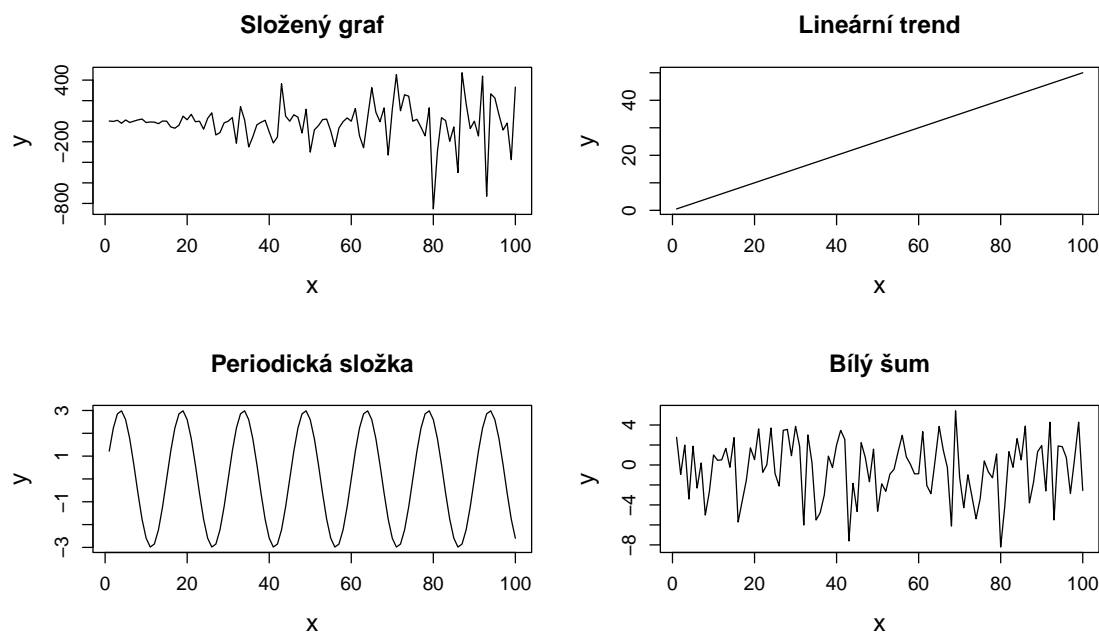
Obrázek 3.4: Aditivní dekompozice

3.1.2 Multiplikativní dekompozice

Při multiplikativní dekompozici předpokládáme rozklad ve tvaru

$$Y_t = Tr_t \cdot Sz_t \cdot C_t \cdot e_t. \quad (3.2)$$

V tomto případě má stejnou jednotku jako původní řada Y_t pouze jedna složka, obvykle ta trendová Tr_t , ostatní složky jsou bezrozměrné a vyjadřují relativní změnu [4].



Obrázek 3.5: Multiplikativní dekompozice

3.2 Boxova-Jenkinsova metodologie

Zatímco dekompozice časových řad vychází z předpokladu nekorelovanosti dat a klade důraz na systematické složky, Boxova-Jenkinsova metodologie se zamerňuje na korelovanost dat a vychází tedy z korelační analýzy [2].

3.3 Lineární dynamické modely

Lineární dynamické modely předpokládají, že vývoj budoucích hodnot dané časové řady závisí jak na minulých hodnotách této řady, tak na hodnotách dalších časových řad. Tyto modely zahrnují například klasické autoregresní modely (AR, ARMA, ARIMA) [7].

3.4 Spektrální analýza

První tři přístupy je možné spojit jedním společným označením a to analýzou v časové doméně. Naproti tomu spektrální analýza se zabývá studiem časových řad ve spektrální doméně, tj. jako nekonečnou kombinací sinových a cosinových funkcí.

Pomocí Fourierovy transformace se řada převede z časové do frekvenční domény a vytvoří se periodogram, ze kterého se dle velikosti amplitud určí významné frekvence. K těmto významným frekvencím se dále hledají pomocí regrese jejich amplitudy a fáze [15].

4 Určení trendu

4.1 Lineární regrese

Tato kapitola čerpá ze skript [1]. Lineární regrese umožňuje zkoumat rozdělení náhodného vektoru s nezávislými složkami, jehož střední hodnota závisí na nezávislých proměnných.

4.1.1 Lineární regresní model plné hodnosti

Rovnice lineárního regresního modelu plné hodnosti je tvaru

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e.$$

Pro n pozorování máme

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nk} \end{pmatrix}, e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

Tedy

$$Y = X\beta + e.$$

Y vysvětlovaná proměnná

$X = (X_1, \dots, X_k)$ vektor vysvětlující proměnné

x_1, \dots, x_k náhodné veličiny vektoru $X = (X_1, \dots, X_k)$

β_0, \dots, β_k parametry regrese

e reziduum

Předpoklady

1. $Ee = 0$, proto $EY = X\beta$
2. $\text{var}(e) = \sigma^2 I$, $\sigma^2 > 0$ je neznámý parametr, pro který platí $De_i = \sigma^2$, $C(e_i, e_j) = 0$, $i \neq j$, tzn. chyby jsou nekorelované a je splněna homoskedasticita

Definice 4.1. Lineární regresní model značíme $\text{LRM}(Y, X\beta, \sigma^2 I)$ a znamená

$$\begin{aligned} Y &= X\beta + e \\ EY &= X\beta \\ \text{var}(Y) &= \text{var}(e) = \sigma^2 I \end{aligned}$$

4.1.2 Metoda nejmenších čtverců

Věta 4.2. Odhad parametru β metodou nejmenších čtverců v $\text{LRM}(Y, X\beta, \sigma^2 I)$ je dán vztahem

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Důkaz. 1.

$$X^T (Y - X\hat{\beta}) = X^T (Y - X(X^T X)^{-1} X^T Y) = 0_{k \times 1}$$

2.

$$\begin{aligned}
(Y - X\beta)^T(Y - X\beta) &= (Y - X\hat{\beta} + X\hat{\beta} - X\beta)^T(Y - X\hat{\beta} + X\hat{\beta} - X\beta) \\
&= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (X\hat{\beta} - X\beta)^T(Y - X\hat{\beta}) + \\
&\quad + (Y - X\hat{\beta})^T(X\hat{\beta} - X\beta) + (X\hat{\beta} - X\beta)^T(X\hat{\beta} - X\beta) \\
&= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + 2(\hat{\beta} - \beta)^T X^T(Y - X\hat{\beta}) + (\hat{\beta} - \beta)^T X^T X(\hat{\beta} - \beta) \\
&\geq (Y - X\hat{\beta})^T(Y - X\hat{\beta}) = \text{RSS}
\end{aligned}$$

3. $X^T X$ je regulární a tedy pozitivně definitní, proto poslední rovnost platí, pokud $\beta = \hat{\beta}$ □

Soustava $X^T X \hat{\beta} = X^T Y$ se nazývá soustava normálních rovnic. Vektor $\hat{Y} = X \hat{\beta} = X(X^T X)^{-1} X^T Y$ lze považovat za minimální aproximaci vektoru Y , kterou lze vytvořit lineární kombinací sloupců matice X .

4.1.3 Prostý lineární regresní model

Jedná se o model tvaru

$$Y = \beta_0 + \beta_1 x + e.$$

Věta 4.3. Platí $E\hat{\beta} = \beta$ a $\text{var}\hat{\beta} = \sigma^2(X^T X)^{-1}$

Důkaz.

$$\begin{aligned}
E\hat{\beta} &= E(X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T EY = (X^T X)^{-1} X^T X \beta = \beta \\
\text{var}\hat{\beta} &= (X^T X)^{-1} X^T (\text{var}Y) X (X^T X)^{-1} = (X^T X)^{-1} X^T (\sigma^2 I) X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}
\end{aligned}$$

□

Věta 4.4. Pro reziduální součet čtverců matice plánu $X = (1, X_1)$ RSS platí

$$\text{RSS} = Y^T Y - \hat{\beta}^T X^T Y.$$

Důkaz.

$$\begin{aligned}
\text{RSS} &= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) = Y^T Y - 2Y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta} = Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X(X^T X)^{-1} X^T Y = \\
&= Y^T Y - \hat{\beta}^T X^T Y
\end{aligned}$$

□

Definice 4.5. Pro naivní model s maticí plánu $X = (1)$ platí

$$\begin{aligned}
\hat{\beta} &= (1^T 1)^{-1} 1^T Y = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}, \hat{Y} = E\hat{Y} = \bar{Y} 1 \\
\text{RSS}_0 &= Y^T (I - \frac{1}{n} \cdot 1 \cdot 1^T) Y = \sum_{i=1}^n (Y_i - \bar{Y})^2 = S_n^2(n-1)
\end{aligned}$$

Poté můžeme lineární model s maticí plánu $X = (1, X_1)$ porovnat s naivním, tj. konstantním modelem s maticí plánu $X = (1)$ a určit koeficient determinace, který je dán vztahem

$$R^2 = 1 - \frac{\text{RSS}}{\text{RSS}_0} \in \langle 0, 1 \rangle,$$

případně korigovaný koeficient determinace

$$R_{adj}^2 = 1 - \frac{n-1}{n-k} \frac{\text{RSS}}{\text{RSS}_0} \in \langle 0, 1 \rangle,$$

Hodnota koeficientu blízká 1 znamená větší regresní závislost.

4.1.4 Obecnější lineární regresní model

Lineární regresní model lze použít i na jiné funkce, než funkce 1. a 0. stupně. Jedná se o model splňující rovnici

$$Y = \beta_0 + \beta_1 f_1(x) + \dots + \beta_k f_k(x) + e. \quad (4.1)$$

Musíme se ale zamyslet nad homoskedasticitou dat. V zásadě jsou tři možnosti.

1. V ideálním případě můžeme předpokládat, že závislost velikosti rozptylu na nezávislé proměnné je stejná jako trendová složka časové řady Tr_t . Tudíž data jsou heteroskedastická před transformací, ale homoskedastická po ní a rezidua bereme po celé časové řadě v jejich neupravené velikosti po transformaci.
2. V praxi se často setkáme s případy, kdy jsou data homoskedastická před transformací a transformováním tudíž vznikne heteroskedasticita. Tento případ řešíme poté váženou regresi, kdy musíme přiřadit reziduíům váhy podle nezávislé proměnné, v našem případě podle času t .
3. Posledním případem je heteroskedasticita s jinou funkční závislostí, než má trendová složka Tr_t . V tomto případě se používají GARCH modely (generalized autoregressive conditional heteroskedasticity, tj. zobecněný autoregresní model s podmíněnou heteroskedasticitou), které jsou tvaru

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

Kde:

- σ_t^2 je rozptyl v čase t
- ω je konstanta
- α_i jsou parametry GARCH modelu pro rezidua ϵ_{t-i}
- ϵ_{t-i} jsou rezidua v čase $t - i$
- β_j jsou parametry GARCH modelu pro rozptyly σ_{t-j}^2
- σ_{t-j}^2 je rozptyl v čase $t - j$
- p je řád autoregrese reziduí
- q je řád autoregrese rozptylů

Konstantní trend

Trend s maticí plánu $X = (1)$, určený rovnicí

$$Tr_t = \beta_0, \quad t = 1, \dots, n.$$

Odhad parametru β_0

$$\sum_{t=1}^n \frac{y_t}{n}$$

Lineární trend

Trend s maticí plánu $X = (1, X_1)$ určený rovnicí

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

Pomocí metody nejmenších čtverců hledáme odhad parametru $\hat{\beta}$ ze soustavy normálních rovnic. Tu můžeme získat buď derivováním vztahu

$$\min_{b_0, b_1} \sum_{t=1}^n (y_t - b_0 - b_1 t)^2$$

podle b_0 a b_1 . Z čehož dostaneme soustavu

$$\begin{aligned} \sum_{t=1}^n y_t - nb_0 - b_1 \sum_{t=1}^n t &= 0 \\ \sum_{t=1}^n ty_t - b_0 \sum_{t=1}^n t - b_1 \sum_{t=1}^n t^2 &= 0, \end{aligned}$$

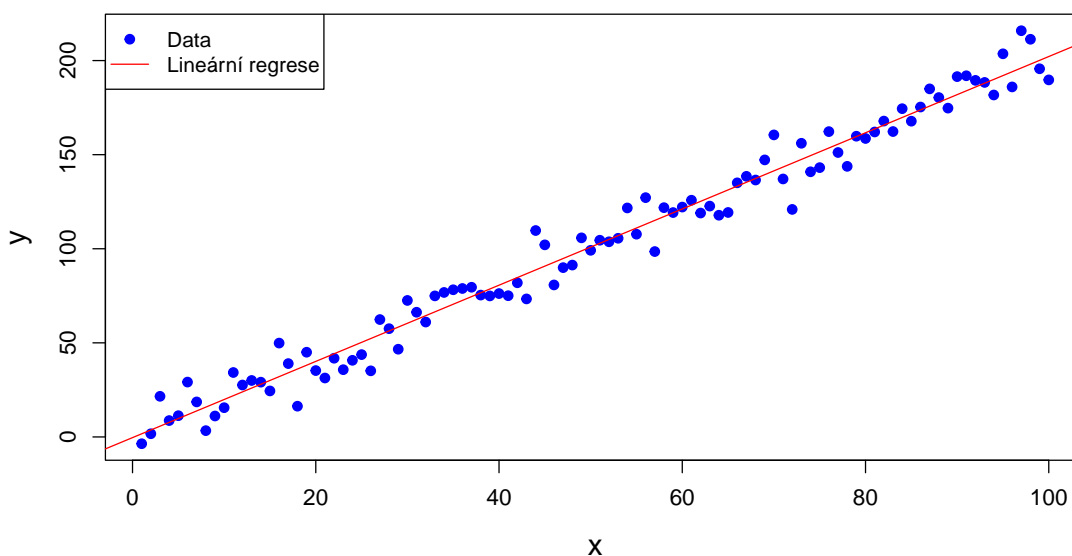
nebo pomocí maticových úprav z matice plánu

$$Y = X\beta + e \Rightarrow (X^T X)\beta = X^T Y \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y.$$

Těmito úpravami dostaneme

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(\begin{pmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \right)^{-1} \cdot \begin{pmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{pmatrix} \cdot \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Na obrázku 4.1 můžeme vidět časovou řadu se silným lineárním trendem.



Obrázek 4.1: časová řada s lineárním trendem

Kvadratický trend

Trend s maticí plánu $X = (1, X_1, X_1^2)$ určený rovnicí

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

Pomocí metody nejmenších čtverců hledáme odhady parametrů $\hat{\beta}$ ze soustavy normálních rovnic. Tu můžeme opět získat buď derivováním vztahu

$$\min_{b_0, b_1} \sum_{t=1}^n (y_t - b_0 - b_1 t - b_2 t^2)^2$$

podle b_0 , b_1 a b_2 . Poté řešíme soustavu

$$\begin{aligned} \sum_{t=1}^n y_t - nb_0 - b_1 \sum_{t=1}^n t - b_2 \sum_{t=1}^n t^2 &= 0, \\ \sum_{t=1}^n ty_t - b_0 \sum_{t=1}^n t - b_1 \sum_{t=1}^n t^2 - b_2 \sum_{t=1}^n t^3 &= 0, \\ \sum_{t=1}^n t^2 y_t - b_0 \sum_{t=1}^n t^2 - b_1 \sum_{t=1}^n t^3 - b_2 \sum_{t=1}^n t^4 &= 0, \end{aligned}$$

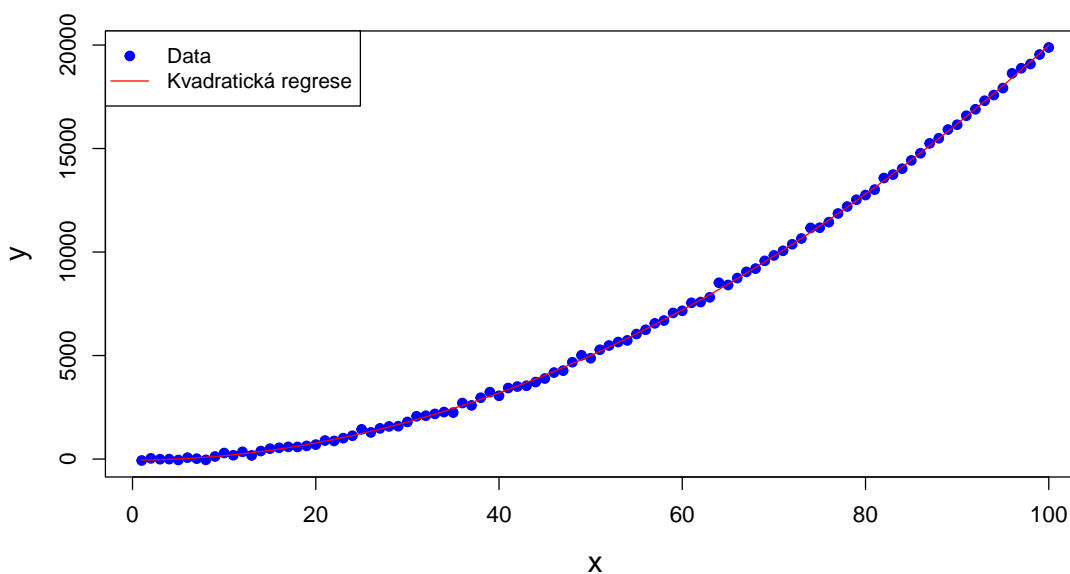
nebo maticovými úpravami, jako v lineárním případě

$$Y = X\beta + e \Rightarrow (X^T X)\beta = X^T Y \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y.$$

Zde výsledná rovnice v maticovém tvaru vypadá takto

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \left(\begin{pmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \\ X_1^2 & \cdots & X_n^2 \end{pmatrix} \begin{pmatrix} 1 & X_1 & X_1^2 \\ \vdots & \vdots & \vdots \\ 1 & X_n & X_n^2 \end{pmatrix} \right)^{-1} \cdot \begin{pmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \\ X_1^2 & \cdots & X_n^2 \end{pmatrix} \cdot \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Na obrázku 4.2 můžeme vidět časovou řadu se silným kvadratickým trendem.



Obrázek 4.2: časová řada s kvadratickým trendem

Exponenciální trend

Rovnice trendu ve tvaru

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

se logaritmováním převede na lineární trend tvaru

$$\ln Tr_t = \ln \alpha + t \ln \beta, \quad t = 1, \dots, n.$$

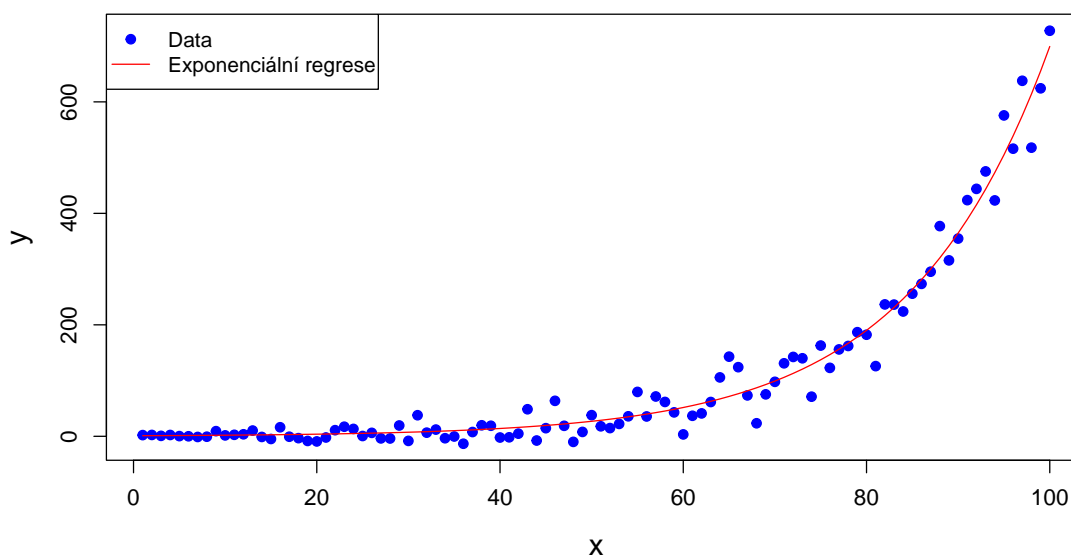
Kvůli logaritmování je nutno použít metodu nejmenších vážených čtverců s váhami v_t . Metoda spočívá v minimalizaci výrazu

$$\sum_{t=1}^n v_t (y_t - \alpha \beta^t)^2$$

Pokud nemáme důvod některá měření preferovat, volíme váhy $v_t = 1$, $t = 1, \dots, n$. Transformované váhy potom jsou $w_t = y_t^2 v_t = y_t^2$ a minimalizujeme výraz

$$\sum_{t=1}^n w_t (\ln y_t - \ln \alpha - t \ln \beta)^2$$

Na obrázku 4.3 můžeme vidět časovou řadu se silným exponenciálním trendem.



Obrázek 4.3: časová řada s exponenciálním trendem

4.1.5 Informativní testy pro volbu trendové křivky

Tabulka 4.1: Informativní testy

Trend	Informativní test
lineární	první diference $y_{t+1} - y_t$ jsou přibližně konstantní
kvadratický	druhé diference $y_{t+2} - 2y_{t+1} + y_t$ jsou přibližně konstantní
exponenciální	podíly sousedních hodnot $\frac{y_{t+1}}{y_t}$ (první diference logaritmů $\log y_{t+1} - \log y_t$) jsou přibližně konstantní

4.2 Adaptivní metody

4.2.1 Jednoduchý klouzavý průměr

Jednoduché klouzavé průměry jsou prosté aritmetické průměry s rovnoměrně rozdělenými váhami. V mé práci se díky svým vlastnostem nedají použít, ale jelikož jsou často využívaným prostředkem, protože jsou jednoduché, tak zde shrnu jejich nejdůležitější specifika, více viz [4].

Jednoduché klouzavé průměry (SMA - simple moving average) mohou být liché délky, pak jsou definovány následovně

$$\hat{y}_t = \frac{1}{2m+1}(y_{t-m} + y_{t-m+1} + \dots + y_{t+m-1} + y_{t+m}),$$

nebo sudé délky. V případě sudé délky by však vyrovnaná hodnota ležela mezi dvěma sousedními hodnotami, proto se používá tzv. *centrování* a vyrovnané hodnoty jsou poté definovány tímto způsobem

$$\hat{y}_t = \frac{1}{4m}(y_{t-m} + 2y_{t-m+1} + \dots + 2y_{t+1-1} + y_{t+m}).$$

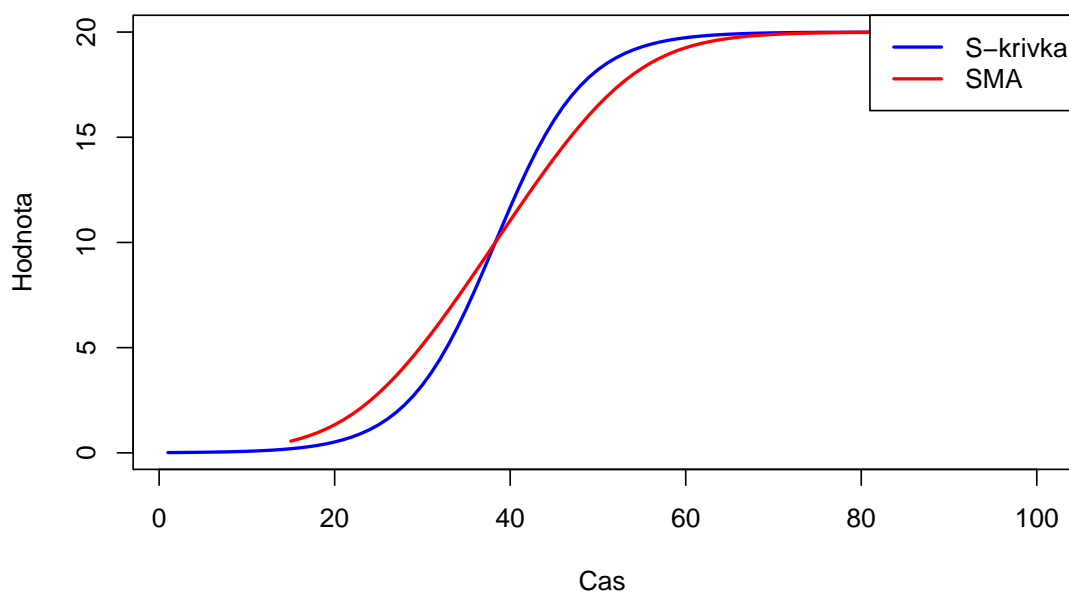
Hlavním požadavkem je, aby klouzavý průměr neměl vliv na trendovou složku Tr_t , bohužel v některých případech tomu tak není, příkladem může být S-křivka, definována následovně

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t}, \quad \beta \in (0, 1), \gamma > 0.$$

Konkrétně v našem případě na 4.4 s těmito parametry

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t} = \frac{20}{1 + 2000 \cdot 0.82^t}$$

a délka klouzevého průměru $n = 29$.



Obrázek 4.4: S-křivka vyhlazená pomocí klouzavého průměru

Vliv vyhlazování pomocí jednoduchých klouzavých průměrů na periodické složky můžeme ukázat na sinusoidě a klouzavém průměru délky $n = 2m + 1$. Využijeme vzorec

$$\frac{1}{2m+1} \sum_{\tau=-m}^m \sin(\omega(t+\tau) + \phi) = \frac{\sin((2m+1)\omega/2)}{(2m+1)\sin(\omega/2)} \sin(t\omega + \phi).$$

Z vzorce vidíme, že nám jednoduchý klouzavý průměr sinusoidu transformuje na sinusoidu se stejnou frekvencí ω i stejnou fází ϕ a amplitudou vynásobenou výrazem

$$\frac{\sin((2m+1)\omega/2)}{(2m+1)\sin(\omega/2)}.$$

Tento výraz je pro nízkou frekvenci ω blízký 1, ale pro velké frekvence se blíží 0. Odtud plyne, že při dlouhodobějších výkyvech se značná část periodické složky zahrne do vyrovnané časové řady, ale kratší výkyvy se nevyrovnejí, odtud lze časovou řadu vyrovnanou pomocí jednoduchých klouzavých průměrů psát ve tvarech

$$Y_t = Tr_t + C_t \quad \text{a} \quad Y_t = Tr_t C_t.$$

Nakonec rezidua přestávají být vlivem klouzavých průměrů nekorelované.

4.2.2 Exponenciální vyrovnání

Tato kapitola pojednává o základních myšlenkách exponenciálního vyrovnání, které v souvislosti se statistikou poprvé uvedl a detailně popsal ve své knize R.G.Brown [3].

Exponenciální vyrovnání dává poslednímu pozorování nejvyšší váhu a tato váha se dále do minulosti exponenciálně snižuje. Pro jeho výpočet máme v zásadě dvě možnosti.

První a obvykle užívanou možností, jak vypočítat exponenciální vyrovnání, je volba délky datového intervalu, který chceme zohlednit N a počítání podle vzorce

$$EMA(t) = Cena(t) \cdot k + EMA(t-1) \cdot (k-1), \text{ kde } k = \frac{2}{(N+1)},$$

délka N určuje míru vyhlazení časové řady.

My se budeme zabývat druhým způsobem výpočtu, který je poněkud složitější, ale praktičtější a vychází z fundamentálního teorému již zmíněného R.G.Browna.

Fundamentální teorém exponenciálního vyrovnání

Uvažujme model reprezentující časovou řadu ve tvaru

$$\xi_t = a_0 + a_1 t + \frac{a_2}{2} t^2 + \dots + \frac{a_n}{n!} t^n.$$

Předpověď budoucích hodnot lze nejlépe vyjádřit jako rozvoj Taylorovy řady v čase t

$$\hat{x}_{t+\tau} = \hat{x}_t^{(0)} + \tau \hat{x}_t^{(1)} + \frac{\tau^2}{2} \hat{x}_t^{(2)} + \dots + \frac{\tau^n}{n!} \hat{x}_t^{(n)},$$

kde $x_t^{(k)}$ je k -tá derivace v čase t a $\hat{x}_t^{(0)}$ je její odhad. Odtud můžeme psát

$$\hat{x}_{t+\tau} = \sum_{k=0}^n \frac{\tau^k \hat{x}_t^{(k)}}{k!} = \sum_{k=0}^n \frac{\tau^k a_k}{k!}$$

pro předpověď pomocí aktuálních odhadů derivací. Vyrovnanou statistiku řádu k definujeme takto:

$$S_t^{[k]}(x) = (1 - \alpha) S_t^{[k-1]}(x) + \alpha S_{t-1}^{[k]}(x).$$

Fundamentální teorém exponenciálního vyrovnání dokazuje, že je možné určit $(n+1)$ koeficientů (derivací) z polynomu řádu n lineární kombinací prvních $(n+1)$ řádů exponenciálního vyrovnání (důkaz viz. str. 133 v [3]).

Z fundamentálního teorému dostaneme po upravení následující:
vektor vyrovnaných statistik S_t

$$S_t = \begin{pmatrix} S_t^{[1]}(x) \\ \vdots \\ S_t^{[n]}(x) \end{pmatrix},$$

a vektor

$$a = \begin{pmatrix} a_0^{[t]} \\ \vdots \\ a_n^{[t]} \end{pmatrix} = \begin{pmatrix} x_t^{[0]}/0! \\ \vdots \\ x_t^{[0]}/n! \end{pmatrix}.$$

Potom fundamentální teorém lze vyjádřit jako

$$S_t = Ma,$$

kde M je matice prvků $M_{ik} = \frac{(-1)^k (1-\alpha)^i}{(i-1)!} \sum_{j=0}^{\infty} j^k \beta^i \frac{(i-1+j)!}{j!}$. Pomocí úprav přes součty nekonečných řad tvaru $\sum_{j=0}^{\infty} j^n \beta^i$ dostaneme matici pro $n = 5$

$$M = \begin{pmatrix} 1 & \frac{-\alpha}{1-\alpha} & \frac{\alpha(1+\alpha)}{(1-\alpha)^2} & \frac{-\alpha(1+4\alpha+\alpha^2)}{(1-\alpha)^3} & \frac{\alpha(1+11\alpha+11\alpha^2+\alpha^3)}{(1-\alpha)^4} \\ 1 & \frac{-2\alpha}{1-\alpha} & \frac{2\alpha(1+2\alpha)}{(1-\alpha)^2} & \frac{-2\alpha(1+7\alpha+4\alpha^2)}{(1-\alpha)^3} & \frac{2\alpha(1+18\alpha+33\alpha^2+8\alpha^3)}{(1-\alpha)^4} \\ 1 & \frac{-3\alpha}{1-\alpha} & \frac{3\alpha(1+3\alpha)}{(1-\alpha)^2} & \frac{-3\alpha(1+10\alpha+9\alpha^2)}{(1-\alpha)^3} & \frac{3\alpha(1+25\alpha+67\alpha^2+27\alpha^3)}{(1-\alpha)^4} \\ 1 & \frac{-4\alpha}{1-\alpha} & \frac{4\alpha(1+4\alpha)}{(1-\alpha)^2} & \frac{-4\alpha(1+13\alpha+16\alpha^2)}{(1-\alpha)^3} & \frac{4\alpha(1+32\alpha+113\alpha^2+64\alpha^3)}{(1-\alpha)^4} \\ 1 & \frac{-5\alpha}{1-\alpha} & \frac{5\alpha(1+5\alpha)}{(1-\alpha)^2} & \frac{-5\alpha(1+16\alpha+25\alpha^2)}{(1-\alpha)^3} & \frac{5\alpha(1+39\alpha+171\alpha^2+125\alpha^3)}{(1-\alpha)^4} \end{pmatrix}$$

Koeficienty a potom spočítáme pomocí inverzní matice jako

$$a = M^{-1} S_t.$$

Jestliže časovou řadu lze reprezentovat polynomem stupně n , potom všechny derivace stupně $(n + 1)$ a vyšší budou identicky rovny nule.

Jednotlivá exponenciální vyrovnání

Časová řada je dána rovnicí $x(t) = \xi(t) + \epsilon(t)$, kde

$$\xi(t) = \sum_{k=1}^n \frac{a_k}{k!} t^k$$

je systematická složka a $\epsilon(t)$ je reziduální složka.

Jednoduché exponenciální vyrovnání

Jednoduché exponenciální vyrovnání předpokládá, že trend časové řady lze v krátkých intervalech považovat za konstantní. Proces je dán rovnicí $\xi(t) = a_0$.

Vyrovnané statistiky :

$$S_t^{[1]}(x) = (1 - \alpha) \cdot x(t) + \alpha \cdot S_{t-1}^{[1]}(x).$$

Počáteční podmínky :

$$S_0^{[1]}(x) = a_0(0).$$

Odhad koeficientů je dán soustavou :

$$\hat{a}_0(t) = 3 \cdot S_t^{[1]}(x).$$

Předpověď v čase T o τ časových kroků je poté dána vzorcem

$$\hat{x}_\tau(T) = \hat{a}_0(T). \quad (4.2)$$

Předpovědí je zde zřejmě konstantní funkce.

Dvojitě exponenciální vyrovnání

Dvojitě exponenciální vyrovnání předpokládá lineární trend časové řady na krátkých intervalech. Proces dán rovnicí $\xi(t) = a_0 + a_1 t$.

Vyrovnané statistiky :

$$\begin{aligned} S_t^{[1]}(x) &= (1 - \alpha) \cdot x(t) + \alpha \cdot S_{t-1}^{[1]}(x) \\ S_t^{[2]}(x) &= (1 - \alpha) \cdot S_t^{[1]}(x) + \alpha \cdot S_{t-1}^{[2]}(x). \end{aligned}$$

Počáteční podmínky :

$$\begin{aligned} S_0^{[1]}(x) &= a_0(0) - \frac{\alpha}{1 - \alpha} \cdot a_1(0) \\ S_0^{[2]}(x) &= a_0(0) - \frac{2 \cdot \alpha}{1 - \alpha} \cdot a_1(0). \end{aligned}$$

Odhad koeficientů je dán soustavou :

$$\begin{aligned} \hat{a}_0(t) &= 2 \cdot S_t^{[1]}(x) - S_t^{[2]}(x) \\ \hat{a}_1(t) &= \frac{1 - \alpha}{\alpha} \left[S_t^{[1]}(x) - S_t^{[2]}(x) \right]. \end{aligned}$$

Předpověď v čase T o τ časových kroků je poté dána vzorcem

$$\hat{x}_\tau(T) = \hat{a}_0(T) + \tau \cdot \hat{a}_1(T). \quad (4.3)$$

Předpověď je v tomto případě lineární funkce.

Trojitě exponenciální vyrovnání

U trojitěho exponenciálního vyrovnání předpokládáme trend v krátkých intervalech kvadratický. Proces $\xi(t) = a_0 + a_1 t + \frac{1}{2} a_2 t^2$.

Vyrovnané statistiky :

$$\begin{aligned} S_t^{[1]}(x) &= (1 - \alpha) \cdot x(t) + \alpha \cdot S_{t-1}^{[1]}(x) \\ S_t^{[2]}(x) &= (1 - \alpha) \cdot S_t^{[1]}(x) + \alpha \cdot S_{t-1}^{[2]}(x) \\ S_t^{[3]}(x) &= (1 - \alpha) \cdot S_t^{[2]}(x) + \alpha \cdot S_{t-1}^{[3]}(x). \end{aligned}$$

Počáteční podmínky :

$$\begin{aligned} S_0^{[1]}(x) &= a_0(0) - \frac{\alpha}{1 - \alpha} \cdot a_1(0) + \frac{\alpha \cdot (2 - \alpha)}{2 \cdot (1 - \alpha)^2} \cdot a_2(0) \\ S_0^{[2]}(x) &= a_0(0) - \frac{2 \cdot \alpha}{1 - \alpha} \cdot a_1(0) + \frac{2 \cdot \alpha \cdot (1 + 2\alpha)}{2 \cdot (1 - \alpha)^2} \cdot a_2(0) \\ S_0^{[3]}(x) &= a_0(0) - \frac{3 \cdot \alpha}{1 - \alpha} \cdot a_1(0) + \frac{3 \cdot \alpha \cdot (1 + 3\alpha)}{2 \cdot (1 - \alpha)^2} \cdot a_2(0). \end{aligned}$$

Odhad koeficientů je dán soustavou :

$$\begin{aligned} \hat{a}_0(t) &= 3 \cdot S_t^{[1]}(x) - 3 \cdot S_t^{[2]}(x) + S_t^{[3]}(x) \\ \hat{a}_1(t) &= \frac{1 - \alpha}{2\alpha^2} \left[(1 + 5\alpha) \cdot S_t^{[1]}(x) - 2(1 + 4\alpha) \cdot S_t^{[2]}(x) + (1 + 3\alpha) \cdot S_t^{[3]}(x) \right] \\ \hat{a}_2(t) &= \frac{1 - \alpha^2}{\alpha^2} \left[(S_t^{[1]}(x) - 2 \cdot S_t^{[2]}(x) + S_t^{[3]}(x)) \right]. \end{aligned}$$

Předpověď v čase T o τ časových kroků je poté dána vzorcem

$$\hat{x}_\tau(T) = \hat{a}_0(T) + \tau \cdot \hat{a}_1(T) + \frac{1}{2} \cdot (\tau^2) \cdot \hat{a}_2(T). \quad (4.4)$$

Předpovědí je kvadratická funkce.

Čtyřnásobné exponenciální vyrovnání

Předpokladem čtyřnásobného exponenciálního vyrovnání je, že je trend v krátkých intervalech polynomem třetího stupně s kladnými koeficienty. Proces dán rovnicí $\xi(t) = a_0 + a_1t + \frac{1}{2}a_2t^2 + \frac{1}{6}a_3t^3$. Vyrovnané statistiky :

$$\begin{aligned} S_t^{[1]}(x) &= (1 - \alpha) \cdot x(t) + \alpha \cdot S_{t-1}^{[1]}(x) \\ S_t^{[2]}(x) &= (1 - \alpha) \cdot S_t^{[1]}(x) + \alpha \cdot S_{t-1}^{[2]}(x) \\ S_t^{[3]}(x) &= (1 - \alpha) \cdot S_t^{[2]}(x) + \alpha \cdot S_{t-1}^{[3]}(x) \\ S_t^{[4]}(x) &= (1 - \alpha) \cdot S_t^{[3]}(x) + \alpha \cdot S_{t-1}^{[4]}(x). \end{aligned}$$

Počáteční podmínky :

$$\begin{aligned} S_0^{[1]}(x) &= a_0(0) - \frac{\alpha}{\alpha - 1}a_1(0) + \frac{\alpha(\alpha + 1)}{2(\alpha - 1)^2}a_2(0) - \frac{\alpha(\alpha^2 + 4\alpha + 1)}{6(\alpha - 1)^3}a_3(0) \\ S_0^{[2]}(x) &= a_0(0) - \frac{2\alpha}{\alpha - 1}a_1(0) + \frac{2\alpha(2\alpha + 1)}{(2\alpha - 1)^2}a_2(0) - \frac{2\alpha(4\alpha^2 + 7\alpha + 1)}{6(\alpha - 1)^3}a_3(0) \\ S_0^{[3]}(x) &= a_0(0) - \frac{3\alpha}{\alpha - 1}a_1(0) + \frac{3\alpha(3\alpha + 1)}{2(\alpha - 1)^2}a_2(0) - \frac{3\alpha(9\alpha^2 + 10\alpha + 1)}{6(\alpha - 1)^3}a_3(0) \\ S_0^{[4]}(x) &= a_0(0) - \frac{4\alpha}{\alpha - 1}a_1(0) + \frac{4\alpha(4\alpha + 1)}{2(\alpha - 1)^2}a_2(0) - \frac{4\alpha(16\alpha^2 + 13\alpha + 1)}{6(\alpha - 1)^3}a_3(0). \end{aligned}$$

Odhad koeficientů je dán soustavou :

$$\begin{aligned} \hat{a}_0(t) &= 4 \cdot S_t^{[1]}(x) - 6 \cdot S_t^{[2]}(x) + 4 \cdot S_t^{[3]}(x) - S_t^{[4]}(x) \\ \hat{a}_1(t) &= \frac{S_t^{[1]}(x) \cdot (-13\alpha^3 + 9\alpha^2 + 3\alpha + 1)}{3\alpha^3} - \frac{S_t^{[2]}(x) \cdot (-19\alpha^3 + 12\alpha^2 + 5\alpha + 2)}{2\alpha^3} \\ &\quad + \frac{S_t^{[3]}(x) \cdot (-7\alpha^3 + 4\alpha^2 + 2\alpha + 1)}{\alpha^3} - \frac{S_t^{[4]}(x) \cdot (-11\alpha^3 + 6\alpha^2 + 3\alpha + 2)}{6\alpha^3} \\ \hat{a}_2(t) &= \frac{S_t^{[1]}(x) \cdot (3\alpha^3 - 5\alpha^2 + \alpha + 1)}{\alpha^3} - \frac{S_t^{[2]}(x) \cdot (8\alpha^3 - 13\alpha^2 + 2\alpha + 3)}{\alpha^3} \\ &\quad + \frac{S_t^{[3]}(x) \cdot (7\alpha^3 - 11\alpha^2 + \alpha + 3)}{\alpha^3} - \frac{S_t^{[4]}(x) \cdot (2\alpha^3 - 3\alpha^2 + 1)}{\alpha^3} \\ \hat{a}_3(t) &= -\frac{S_t^{[1]}(x) \cdot (\alpha^3 - 3\alpha^2 + 3\alpha - 1)}{\alpha^3} + \frac{3 \cdot S_t^{[2]}(x) \cdot (\alpha^3 - 3\alpha^2 + 3\alpha - 1)}{\alpha^3} \\ &\quad - \frac{3 \cdot S_t^{[3]}(x) \cdot (\alpha^3 - 3\alpha^2 + 3\alpha - 1)}{\alpha^3} + \frac{S_t^{[4]}(x) \cdot (\alpha^3 - 3\alpha^2 + 3\alpha - 1)}{\alpha^3}. \end{aligned}$$

Předpověď v čase T o τ časových kroků je poté dána vzorcem

$$\hat{x}_\tau(T) = \hat{a}_0(T) + \tau \cdot \hat{a}_1(T) + \frac{1}{2} \cdot (\tau^2) \cdot \hat{a}_2(T) + \frac{1}{6} \cdot (\tau^3) \cdot \hat{a}_3(T). \quad (4.5)$$

Předpovědi čtyřnásobného vyrovnání je funkce třetího stupně s kladnými koeficienty.

Z rovnic předpovědi je zřejmé, že není reálné použití vyšších řádů exponenciálního vyrovnání pro předpovědi o více, než jen několik málo kroků do budoucnosti, tuto skutečnost budu později demonstrovat v praktické části.

Volba parametru α

Z praktických zkušeností vyplývá, že při volbě konstanty α má smysl uvažovat pouze interval $0,7 \leq \alpha < 1$ [4]. Menší hodnoty parametru α totiž dávají větší váhu poslednímu pozorování a menší váhu exponenciálnímu vyrovnání, čímž se snižuje účinek vyrovnání.

Při hledání parametru α nicméně musíme testovat i hodnoty z intervalu $(0; 0,7)$, protože, pokud nám vyjde α v tomto intervalu, tak víme, že námi použitý model neodpovídá modelované situaci. Toto může být způsobeno buď špatným předpokladem na lokální aproximaci trendu, nebo vysokou mírou autokorelace dat.

Já se ve své práci zabývám pouze první možností a to nesprávným odhadem lokální trendové funkce. Řešení tohoto problému spočívá v hledání a testování exponenciálních vyrovnání vyšších řádů. V mém případě se parametr α dostal do intervalu $(0,7; 1)$ pro každou zkoumanou časovou řadu vždy až pro čtyřnásobné exponenciální vyrovnání.

Mnou použité experimentální hledání vhodné hodnoty parametru α spočívá ve vyrovnávání časové řady s různými hodnotami α a porovnávání součtů kvadratických chyb reziduí.

5 Určení cyklické složky

5.1 Spektrální analýza

Zdrojem pro tuto kapitolu byla zejména kniha [15].

Zatímco předchozí kapitola zabývající se trendovou složkou časové řady se zabývala daty v časové doméně, spektrální analýza se zabývá daty v doméně frekvenční.

5.1.1 Cyklické chování a periodicitá

Definice 5.1 (Periodický proces). Jako periodický proces budeme uvažovat funkci odpovídající rovnici

$$x_t = A \cos(2\pi\omega t + \varphi),$$

kde $t = 0, \pm 1, \pm 2, \dots$, ω je frekvenční index definovaný v cyklech za jednotku času, A je amplituda funkce a φ je fáze. Pro lepší manipulaci přepíšeme rovnici pomocí trigonometrické identity

$$\cos(\alpha \pm \beta) = \cos(\alpha) \cos(\beta) \mp \sin(\alpha) \sin(\beta)$$

na

$$x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t), \quad (5.1)$$

kde $U_1 = A \cos \varphi$ a $U_2 = -A \sin \varphi$ jsou náhodné veličiny s normálním rozdělením. Amplituda je potom $A = \sqrt{U_1^2 + U_2^2}$ a fáze $\varphi = \arctan(-U_2/U_1)$. Z toho poté vyplývá, že A a φ jsou nezávislé náhodné veličiny, kde A^2 má rozložení χ^2 se dvěma stupni volnosti a φ má rovnoměrné rozdělení na intervalu $(-\pi, \pi)$, pak U_1 a U_2 jsou nezávislé normální náhodné veličiny.

Variabilita je v takové časové řadě pak způsobena náhodnou amplitudou A a fází φ .

Definice 5.2 (Autokovariační funkce). Předpokládáme-li, že U_1 a U_2 jsou nekorelované náhodné veličiny s $E(U_1) = E(U_2) = 0$ a $D(U_1) = D(U_2)$, pak $E(x_t) = 0$ a, položíme-li $c_t = \cos(2\pi\omega t)$ a $s_t = \sin(2\pi\omega t)$, autokovariační funkce je potom

$$\begin{aligned} \gamma_x(h) &= \text{cov}(x_{t+h}, x_t) = \text{cov}(U_1 c_{t+h} + U_2 s_{t+h}, U_1 c_t + U_2 s_t) \\ &= \text{cov}(U_1 c_{t+h}, U_1 c_t) + \text{cov}(U_1 c_{t+h}, U_2 s_t) + \text{cov}(U_2 s_{t+h}, U_1 c_t) + \text{cov}(U_2 s_{t+h}, U_2 s_t) \\ &= \sigma^2 c_{t+h} c_t + 0 + 0 + \sigma^2 s_{t+h} s_t = |\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta| = \\ &= \sigma^2 \cos(2\pi\omega h). \end{aligned}$$

Pro $h = 0$ dostáváme

$$\gamma_x(0) = \text{var}(x_t) = \sigma^2.$$

Umožníme-li více součet více periodických procesů, dostáváme zobecnění vztahu 5.1

$$x_t = \sum_{k=1}^q U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t), \quad (5.2)$$

kde U_{k1}, U_{k2} , pro $k = 1, \dots, q$, jsou nekorelované náhodné veličiny s nulovou střední hodnotou a rozptyly σ_k^2 a ω_k jsou po dvou různé periody. To znamená, že proces 5.2 je součet nekorelovaných

procesů s rozptyly σ_k^2 a frekvencemi ω_k . Poté můžeme odvodit zobecnění autokovariační funkce

$$\begin{aligned}
 \gamma_x(h) &= \sum_{k=1}^q \text{cov}(x_{k(t+h)}, x_{kt}) \\
 &= \sum_{k=1}^q \text{cov}(U_{k1}c_{k(t+h)} + U_{k2}s_{k(t+h)}, U_{k1}c_{kt} + U_{k2}s_{kt}) \\
 &= \sum_{k=1}^q \text{cov}((U_{k1}c_{k(t+h)}, U_{k1}c_{kt}) + \text{cov}(U_{k1}c_{k(t+h)}, U_{k2}s_{kt}) \\
 &\quad + \text{cov}(U_{k2}s_{k(t+h)}, U_{k1}c_{kt}) + \text{cov}(U_{k2}s_{k(t+h)}, U_{k2}s_{kt})) \\
 &= \sum_{k=1}^q (\sigma^2 c_{k(t+h)}c_{kt} + 0 + 0 + \sigma^2 s_{k(t+h)}s_{kt}) \\
 &= \sum_{k=1}^q \sigma_k^2 \cos(2\pi\omega_k h).
 \end{aligned}$$

Vidíme, že zobecněná autokovariační funkce je součet periodických složek s váhami úměrnými rozptylům σ_k^2 . Odtud pro $h = 0$ dostáváme

$$\gamma_x(0) = \text{var}(x_t) = \sum_{k=1}^q \sigma_k^2.$$

Takže x_t je stacionární proces s nulovou střední hodnotou a celkovou variancí danou součtem rozptylů jednotlivých složek.

Pokud pozorujeme $U_{k1} = a_k$ a $U_{k2} = b_k$ pro $k = 1, \dots, q$, potom odhad k -té složky rozptylu, σ_k^2 , z $\text{var}(x_t)$ by byl výběrový rozptyl $S_k^2 = a_k^2 + b_k^2$. Dále odhad rozptylu pro $h = 0$ je dán jako součet výběrových rozptylů

$$\hat{\gamma}_x(0) = \hat{\text{var}}(x_t) = \sum_{k=1}^q (a_k^2 + b_k^2).$$

Pro každou časovou řadu, kde n je liché, můžeme psát

$$x_t = a_0 + \sum_{j=1}^{\frac{n-1}{2}} a_j \cos\left(\frac{2\pi t j}{n}\right) + b_j \sin\left(\frac{2\pi t j}{n}\right), \quad (5.3)$$

pokud je n sudé, tak se rovnost změní přidáním $a_{\frac{n}{2}} \cos(2\pi t \frac{1}{2}) = a_{\frac{n}{2}} (-1)^t$ na

$$x_t = a_{\frac{n}{2}} (-1)^t + \sum_{j=1}^{\frac{n-1}{2}} a_j \cos\left(\frac{2\pi t j}{n}\right) + b_j \sin\left(\frac{2\pi t j}{n}\right). \quad (5.4)$$

Vyjádření 5.3 a 5.4 je přesné pro každou uvažovanou řadu, potom 5.2 lze považovat za jeho aproximaci předpokládající, že značné množství koeficientů v 5.2 bude blízkých nule. Regresní koeficienty a_j a b_j jsou tvaru $\sum_{t=1}^n x_t z_{tj} / \sum_{t=1}^n z_{tj}^2$, kde z_{tj} je buď $\cos(2\pi t j/n)$, nebo $\sin(2\pi t j/n)$ [15], užitím kořenů komplexní jednotky můžeme poté koeficienty zapsat jako

$$a_j = \frac{2}{n} \sum_{t=1}^n x_t \cos\left(\frac{2\pi t j}{n}\right), \quad b_j = \frac{2}{n} \sum_{t=1}^n x_t \sin\left(\frac{2\pi t j}{n}\right).$$

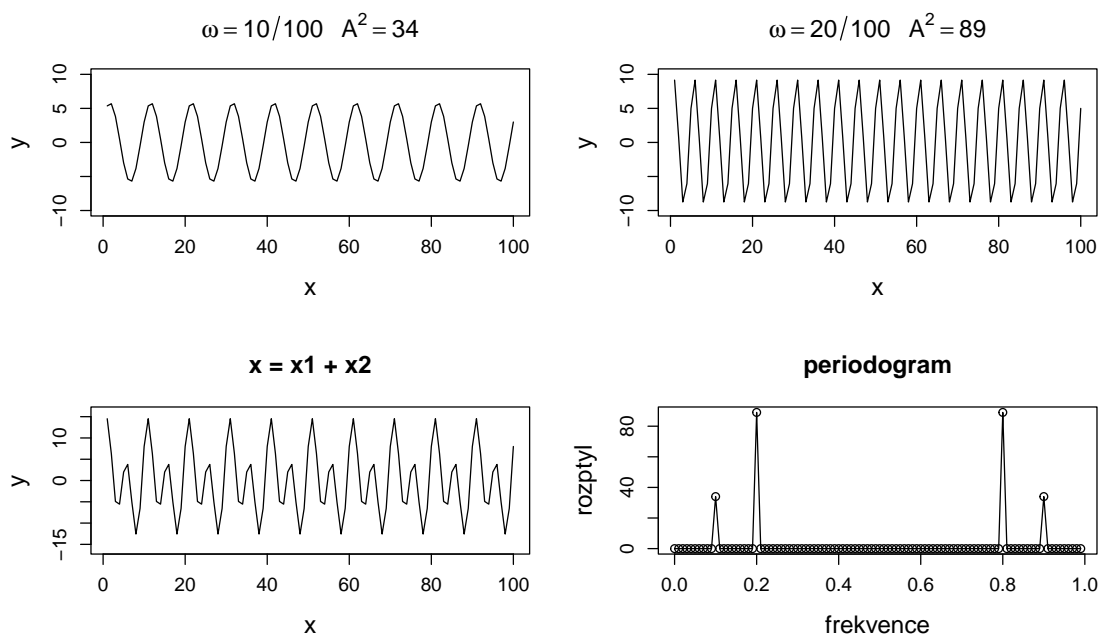
Škálovaný periodogram je poté definován jako

$$P\left(\frac{j}{n}\right) = a_j^2 + b_j^2.$$

V takto definovaném periodogramu můžeme vidět, které frekvence jsou významné a které nikoliv podle velikosti hodnot $P(j/n)$. Periodogram ukazuje výběrový rozptyl každé frekvence $\omega_j = \frac{j}{n}$ a je odhadem σ_j^2 odpovídající frekvenci $\omega = j/n$, nazývané Fourierovy, nebo fundamentální frekvence. Poprvé byl použit roku 1998 A. Schusterem¹ v [14].

Poznámka. Při výpočtu se používají algoritmy diskrétní Fourierovy transformace (DFT) [15] a rychlé Fourierovy transformace (FFT) [5], které významně zrychlí výpočet,

Na obr.5.1 můžeme vlevo dole vidět řadu $x_t = x_{t1} + x_{t2}$, složenou z řady $x_{t1} = 3 \cos(2\pi t \frac{10}{100}) + 5 \sin(2\pi t \frac{10}{100})$, umístěné vlevo nahoře a řady $x_{t2} = 5 \cos(2\pi t \frac{20}{100}) + 8 \sin(2\pi t \frac{20}{100})$, umístěné v pravo nahoře. V pravém dolním rohu poté vidíme škálovaný periodogram řady x_t , ze kterého je patrné, že významné jsou právě periody 0,1 a 0,2, dle očekávání. Dále si můžeme všimnout, že je tu symetrie podle frekvence 0,5, jedná se o tzv. Nyquistovu frekvenci, což je nejvyšší frekvence, jakou můžeme vidět v diskrétním vzorku dat.



Obrázek 5.1: Součet periodických komponent a jejich periodogram

5.1.2 Spektrální hustota

Spektrální hustota je základním pilířem spektrální analýzy. Teorém spektrální reprezentace dat a důkaz, že stacionární časovou řadu lze rozložit na její periodické komponenty úměrné jejich rozptylům je možno najít v dodatku C (Appendix C) v [15].

Uvažujme periodický stacionární náhodný proces z rovnice 5.1 s pevně danou frekvencí ω_0

$$x_t = U_1 \cos(2\pi\omega_0 t) + U_2 \sin(2\pi\omega_0 t), \quad (5.5)$$

¹Sir Arthur Schuster, 1851 - 1934, německo-britský fyzik

kde U_1 a U_2 jsou nekorelované náhodné veličiny se střední hodnotou 0 a shodným rozptylem σ^2 . Počet period potřebných k dokončení jednoho cyklu je přesně $1/\omega$ pro $t = 0, \pm 1, \pm 2, \dots$ a varianci reprezentujeme jako

$$\gamma(h) = \sigma^2 \cos(2\pi\omega_0 h) = \frac{\sigma^2}{2} e^{-2\pi i\omega_0 h} + \frac{\sigma^2}{2} e^{2\pi i\omega_0 h} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} dF(\omega).$$

Poté užitím Riemann-Stieltjes integrace [8] dostáváme

$$F(\omega) = \begin{cases} 0 & \text{pro } \omega < -\omega_0, \\ \frac{\sigma^2}{2} & \text{pro } -\omega_0 \leq \omega < \omega_0, \\ \sigma^2 & \text{pro } \omega \geq \omega_0. \end{cases}$$

Funkce $F(\omega)$ poté funguje jako kumulativní distribuční funkce náhodné diskrétní proměnné s tím rozdílem, že $F(\infty) = \sigma^2$. $F(\omega)$ nazýváme spektrální distribuční funkce a $F(\infty)$ jako celkový rozptyl procesu x_t .

Reprezentace [5.5] pro stacionární řadu existuje vždy.

Věta 5.3 (Spektrální reprezentace autokovariační funkce). *Jestliže x_t je stacionární řada s autokovariancí $\gamma(h) = \text{cov}(x_{t+h}, x_t)$, pak existuje jediná monotónní rostoucí funkce $F(\omega)$ s $F(-\infty) = F(-1/2) = 0$ a $F(\infty) = F(1/2) = \gamma(0)$ taková, že*

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} dF(\omega).$$

Tuto funkci $F(\omega)$ nazýváme funkcí spektrální distribuce.

Věta 5.4. *Jestliže autokovariační funkce $\gamma(h)$ splňuje*

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

potom lze vyjádřit

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} f(\omega) d\omega, \quad \text{pro } h = 0, \pm 1, \pm 2, \dots, \quad (5.6)$$

jako inverzní transformaci spektrální hustoty

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h}, \quad \text{kde } \omega \in \langle -1/2; 1/2 \rangle \quad (5.7)$$

Z rovnice [5.7] vyplývá, že se jedná o sudou funkci, tj. $f(\omega) = f(-\omega)$, což vysvětluje symetrii v našem periodogramu, proto se někdy vykresluje pouze první polovina periodogramu a grafů spektrální hustoty.

Pokud položíme $h = 0$ v [5.6], dostaneme

$$\gamma(0) = \text{var}(x_t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega) d\omega, \quad (5.8)$$

který udává celkovou varianci jako integrál spektrální hustoty přes všechny frekvence.

Spektrální reprezentace stacionárního procesu

Jestliže x_t je stacionární proces s nulovou střední hodnotou a spektrální distribucí $F(\omega)$, potom existuje komplexní stochastický proces $Z(\omega)$ na intervalu $\omega \in \langle -1/2; 1/2 \rangle$ se stacionárními nekorelovanými a nepřekrývajícími se přírůstky takový, že x_t lze vyjádřit jako stochastický integrál [8]

$$x_t = \int_{-1/2}^{1/2} e^{2\pi i \omega t} dZ(\omega),$$

kde pro $-1/2 \geq \omega_1 \geq \omega_2 \geq 1/2$,

$$\text{var}\{Z(\omega_2) - Z(\omega_1)\} = F(\omega_2) - F(\omega_1).$$

5.1.3 Periodogram a diskrétní Fourierova transformace

Definice 5.5 (Diskrétní Fourierova transformace). Pro data x_1, \dots, x_n definujeme diskrétní Fourierovu transformaci (DFT) jako

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t e^{-2\pi i \omega_j t}, \quad \text{pro } j = 0, 1, \dots, n-1,$$

kde frekvence $\omega_j = j/n$ se nazývají Fourierovy, nebo fundamentální frekvence.

Pokud je n složeno z mnoha činitelů, pak lze DFT počítat pomocí zrychlené Fourierovy transformace (FFT), viz např. [5], nebo [6].

Poznámka. FFT algoritmus byl poprvé vynalezen německým matematikem J. C. F. Gaussem v roce 1805 při počítání drah asteroidů Juno a Palas.

Věta 5.6 (Inverzní diskrétní Fourierova transformace).

$$x_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t}, \quad \text{pro } t = 1, \dots, n$$

Definice 5.7 (Periodogram). Pro daná x_1, \dots, x_n definujeme periodogram jako

$$I(\omega_j) = |d(\omega_j)|^2, \quad \text{pro } j = 0, 1, 2, \dots, n-1$$

Poznámka. $I(0) = n\bar{x}^2$, kde \bar{x} je výběrový průměr. Dále $\sum_{t=1}^n \exp(-2\pi i t(j/n)) = 0$ pro $j \neq 0$. Odtud můžeme DFT psát jako

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n (x_t - \bar{x}) e^{-2\pi i \omega_j t} \quad (5.9)$$

pro $j \neq 0$. Odtud poté

$$\begin{aligned} I(\omega_j) &= |d(\omega_j)|^2 = n^{-1} \sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j (t-s)} \\ &= n^{-1} \sum_{h=1-n}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h} \\ &= \sum_{h=1-n}^{n-1} \hat{y}(h) e^{-2\pi i \omega_j h} \end{aligned}$$

pro $j \neq 0$, kde $h = t - s$. Z tohoto pohledu je periodogram $I(\omega_j)$ výběrovou verzí $f(\omega)$ z 5.7.

Definice 5.8 (cosinová a sinová transformace). Pro daná x_1, \dots, x_n definujeme cosinovou transformaci jako

$$d_c(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \cos(2\pi\omega_j t)$$

a sinovou transformaci jako

$$d_s(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \sin(2\pi\omega_j t),$$

kde $\omega_j = j/n$ pro $j = 0, 1, \dots, n-1$.

Poznámka. $d(\omega_j) = d_c(\omega_j) - id_s(\omega_j)$ odtud

$$I(\omega_j) = d_c^2(\omega_j) + d_s^2(\omega_j)$$

Nyní v 5.9 nahradíme \bar{x} za μ a píšeme

$$I(\omega_j) = n^{-1} \sum_{h=1-n}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \mu)(x_t - \mu) e^{-2\pi i \omega_j h},$$

kde ω_j je nenulová fundamentální frekvence. Odtud dostáváme

$$E [I(\omega_j)] = \sum_{h=1-n}^{n-1} \left(\frac{n-|h|}{n} \right) \gamma(h) e^{-2\pi i \omega_j h}. \quad (5.10)$$

Pro dané nenulové ω vybereme posloupnost fundamentálních frekvencí $\omega_{j;n} \rightarrow \omega$, pro kterou z 5.10 vyplývá, když $n \rightarrow \infty$

$$E [I(\omega_{j;n})] \rightarrow f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i h \omega}.$$

Při absolutní sumovatelnosti $\gamma(h)$ je spektrální hustota dlouhodobým průměrem periodogramu. Pokud autokovariační funkce splňuje

$$\Theta = \sum_{h=-\infty}^{\infty} |h| |\gamma(h)| < \infty, \quad (5.11)$$

můžeme odvodit další asymptotické vlastnosti.

$$\text{cov}[d_c(\omega_j), d_c(\omega_k)] = n^{-1} \sum_{s=1}^n \sum_{t=1}^n \gamma(s-t) \cos(2\pi\omega_j s) \cos(2\pi\omega_k t)$$

$$\text{cov}[d_c(\omega_j), d_s(\omega_k)] = n^{-1} \sum_{s=1}^n \sum_{t=1}^n \gamma(s-t) \cos(2\pi\omega_j s) \sin(2\pi\omega_k t)$$

$$\text{cov}[d_s(\omega_j), d_s(\omega_k)] = n^{-1} \sum_{s=1}^n \sum_{t=1}^n \gamma(s-t) \sin(2\pi\omega_j s) \sin(2\pi\omega_k t)$$

Speciálně pro $\omega_j, \omega_k \neq 0$, nebo $1/2$,

$$\text{cov}[(d_c(\omega_j), d_c(\omega_k))] = \begin{cases} f(\omega_j)/2 + \epsilon_n & \text{pro } \omega_j = \omega_k, \\ \epsilon_n & \text{pro } \omega_j \neq \omega_k, \end{cases}$$

$$\text{cov}[(d_s(\omega_j), d_s(\omega_k))] = \begin{cases} f(\omega_j)/2 + \epsilon_n & \text{pro } \omega_j = \omega_k, \\ \epsilon_n & \text{pro } \omega_j \neq \omega_k \end{cases}$$

a

$$\text{cov}[d_c(\omega_j), d_s(\omega_k)] = \epsilon_n,$$

kde chybu ϵ_n v aproximacích lze ohraničit,

$$|\epsilon_n| \leq \Theta/n,$$

kde Θ je dána vztahem 5.11. Nakonec pro případ $\omega_j, \omega_k = 0$, nebo $1/2$ dostaneme

$$\text{cov}[(d_c(\omega_j), d_c(\omega_k))] = \begin{cases} f(\omega_j) + \epsilon_n & \text{pro } \omega_j = \omega_k, \\ \epsilon_n & \text{pro } \omega_j \neq \omega_k, \end{cases}$$

a

$$\text{cov}[d_c(\omega_j), d_s(\omega_k)] = \epsilon_n,$$

$$d_s(0) = d_s(1/2) = 0.$$

Věta 5.9 (Rozdělení periodogramu). *Jestliže*

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

kde w_t jsou nezávislé stejně rozdělené veličiny (označíme iid) s rozdělením $N(0, \sigma_w^2)$ a platí 5.11, pak pro libovolný soubor m různých frekvencí $\omega_j \in (0, 1/2)$ s posloupnostmi $\omega_{j:n} \rightarrow \omega_k$

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} \text{iid } \chi_2^2,$$

za předpokladu, že $f(\omega_j) > 0$, pro $j = 1, \dots, m$.

Poznámka. Periodogram jako odhad významných frekvencí je náchylný k velkým nejistotám a důležitou problematikou, zvláště pro vysoké hodnoty n , je zmenšování jeho variance. Jedná se o složitou problematiku, kterou jsem ve své práci řešil pouze jednoduchým filtrem, tudíž zde pouze uvádím svůj zdroj, kde se tato problematika řeší hlouběji [15].

6 Praktická část

V této části práce aplikuji znalosti získané předchozím studiem na vybrané časové řady a prezentuji výsledky tohoto přístupu v praxi.

V první řadě si musíme uvědomit, že cenové pohyby na trhu a vývoj těchto časových řad je komplexní problém a námi zvolená čistě technická analýza nedokáže vysvětlit veškeré pohyby na trhu. Mezi tyto, technickou analýzou nevysvětlitelné, jevy patří například velké cenové pohyby na cenách ropy způsobené politickými změnami ve světě, nebo například z poslední doby pandemie Covid-19, která značně ovlivnila ceny akcií farmaceutických firem.

Takto významně ovlivněným časovým řadám jsem se při této práci vyhýbal, přestože exponenciální vyrovnání dokáže tyto změny vyhladit, pokud se nestanou bezprostředně před predikcí. Nicméně to není případ lineární regrese, se kterou jsem exponenciální vyrovnání porovnával a rovněž to neznámá, že se v časových řadách neobjevují výkyvy, které bychom mohli lépe vysvětlit fundamentální analýzou místo pouze analýzou technickou.

Jak jsem již zmínil, jednou motivací pro implementaci lineární regrese bylo porovnání s exponenciálním vyrovnáním, ale zásadnějším důvodem je fakt, že lineární regrese umožňuje větší vliv periodických složek a narozdíl od exponenciálního vyrovnání ho můžeme používat i na předpovědi delších časových úseků, na které se exponenciální vyrovnání, zejména vyrovnání vyšší řádů, již nehodí.

S ohledem na to, že bylo mým cílem tyto dva přístupy přímo porovnávat mezi sebou, tak zde použiji pouze předpovědi o 10 časových kroců dopředu, v případě, který zde budu uvádět, se jedná o 10 dní, jelikož budu celou problematiku prezentovat na denních grafech. Tento přístup lze samozřejmě použít i na jiné časové intervaly, já jsem ho sám používal na 15 minutové, hodinové i jiné grafy, nicméně tyto grafy se již nehodí na prezentování bez použití interaktivních grafů, protože nelze tyto grafy rozumně prezentovat v celé své délce ve formátu této práce. Nicméně stále zde platí, že exponenciální vyrovnání vyšších řádů se hodí jen na předpovědi o několik málo kroků dopředu, což může být u 15 minutových grafů jen pár hodin, a proto jsem se rozhodl vše prezentovat na denních grafech. Dále bych zmínil, že analýzu provádím pouze na uzavíracích cenách (close), protože ty nás při předpovědi zajímají nejvíce.

Své výsledky budu prezentovat na grafech 3 různých akcií, které jsem zde vložil pomocí obrázků, ale pro lepší přehled jsou v přílohách grafy těchto vyrovnání s předpovědí v podobě interaktivních html grafů. Jedná se o firmy Admiral Group plc, zabývající se finančními službami, dále akciemi britské zbrojní firmy BAE systems a nakonec společnosti Rolls Royce. Tyto tři grafy poměrně dobře ukazují, k jakým situacím obvykle dochází a tudíž nejsou výběrem nejlepších možných scénářů.

6.1 Implementace

Své výpočty a grafické znázornění výsledků jsem programoval v jazyce Python 16 za použití knihoven Numpy 9, Pandas 10, Plotly 11 a Scipy 13. Zde bych zejména zmínil výpočet lineární regrese a periodogramu pomocí funkcí z knihovny Numpy a jejího FFT algoritmu a následnou regresi těchto periodických složek pomocí funkce `curve_fit` z balíčku Scipy. Exponenciální vyrovnání jsem programoval dle vzorců v kapitole čtyři.

6.2 Akcie Rolls Royce

6.2.1 Výchozí data

První časovou řadou, kterou zde budu prezentovat jsou akcie Rolls-Royce Holdings plc. Na obrázku 6.1 můžeme vidět nejdříve graf za rok 2023. Virtuálně se nacházíme v situaci, kdy známe tato

data a chceme předpovědět průběh na dalších 10 obchodovacích dní dopředu.



Obrázek 6.1: akcie Rolls Royce 1 rok

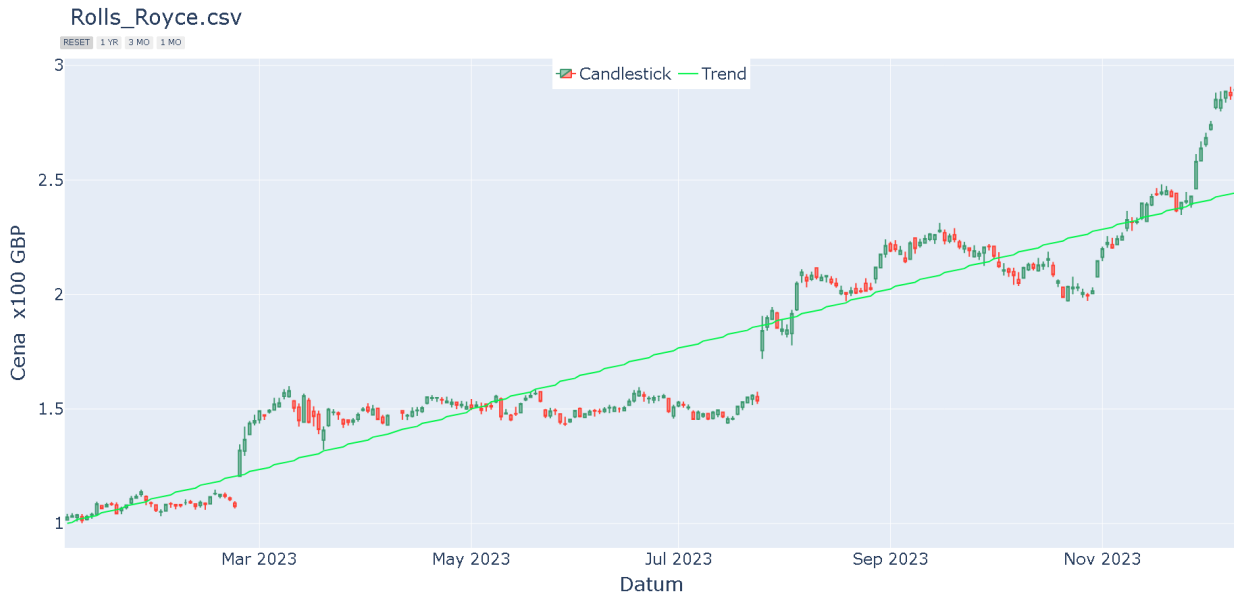
Jelikož je mým cílem ukázat, kdy se který model vyrovnání ukázal jako nejlepší, tak zde vykreslím i graf 6.2, který již ukazuje 10 námi predikovaných hodnot, čtenář se tedy může pokusit odhadnout, který typ vyrovnání by mohl vyjít nejlépe.



Obrázek 6.2: akcie Rolls Royce s predikovaným obdobím

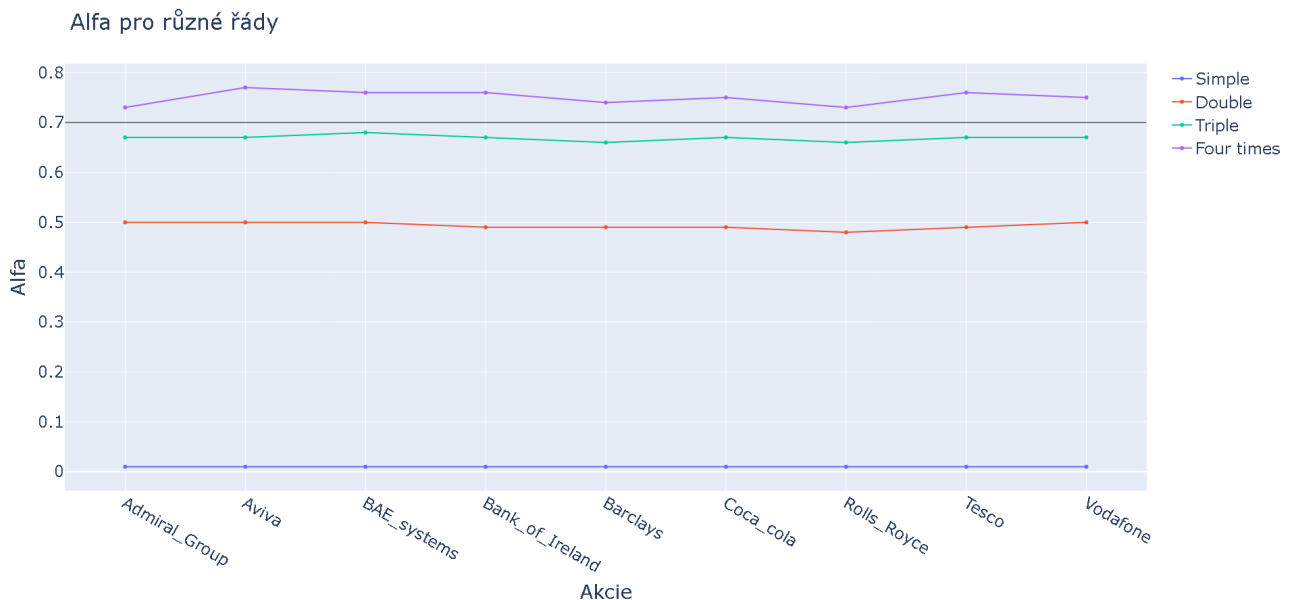
6.2.2 Výpočet trendu

Nejdříve uděláme na známých datech lineární regresi 6.3. Zde můžeme vidět, že výkyvy jsou stále oproti regresní přímce poměrně značné, to ale na druhou stranu umožní větší vliv periodických složek, které budeme hledat spektrální analýzou.



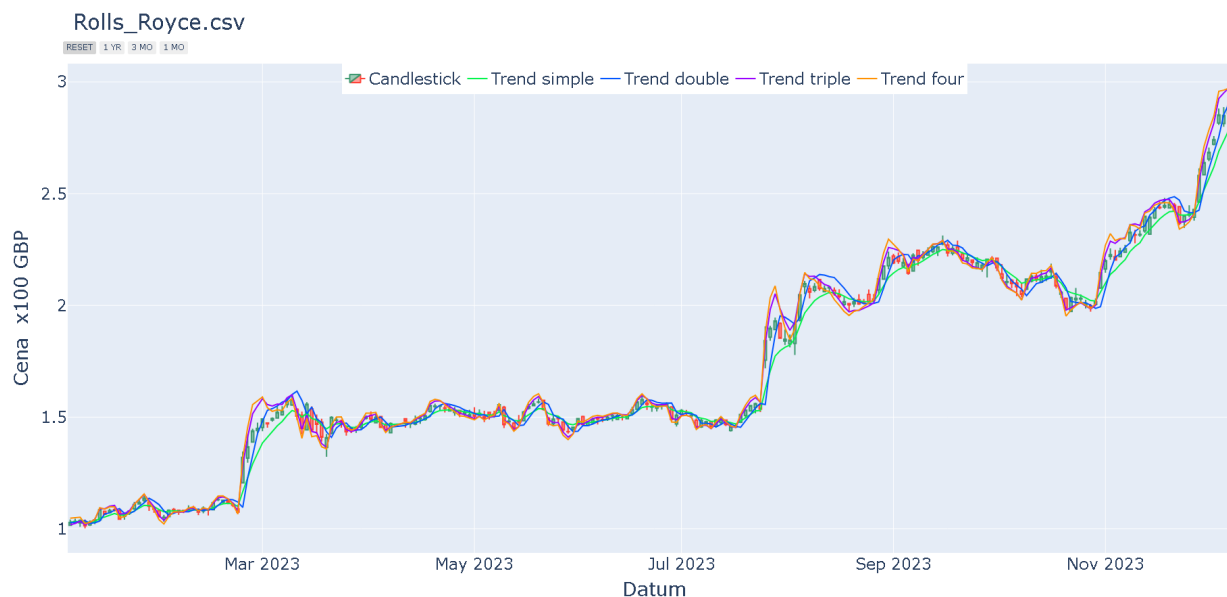
Obrázek 6.3: lineární regrese Rolls Royce

Dále vypočítáme na stejných datech exponenciální vyrovnání 6.5. U exponenciálního vyrovnání musíme mít na paměti podmínky zmíněné v teoretické části o volbě optimálního parametru 4.2.2, tzn. optimální $\alpha \geq 0.7$. To zjistíme experimentálně tak, že spočítáme tzv. SSE (sum of squared errors), tj. součet čtverců reziduí, pro každou hodnotu α zvlášť, v mém případě jsem zvyšoval hodnotu po 0.01, a tyto hodnoty porovnáme. V grafu níže 6.4 můžeme již vidět výsledné optimální hodnoty parametru α .



Obrázek 6.4: vypočtené parametry α

Z tohoto grafu 6.4 vyplývá, že tato podmínka je přirozeně splněna pouze pro čtyřnásobné exponenciální vyrovnání, pro ostatní vyrovnání jsou nalezené hodnoty nižší a musejí být manuálně nastaveny na dolní hranici parametru, tj. $\alpha = 0.7$, to děláme proto, aby vyrovnané statistiky měly převážný vliv na hodnotu vyrovnání a poté hodnotu predikce.



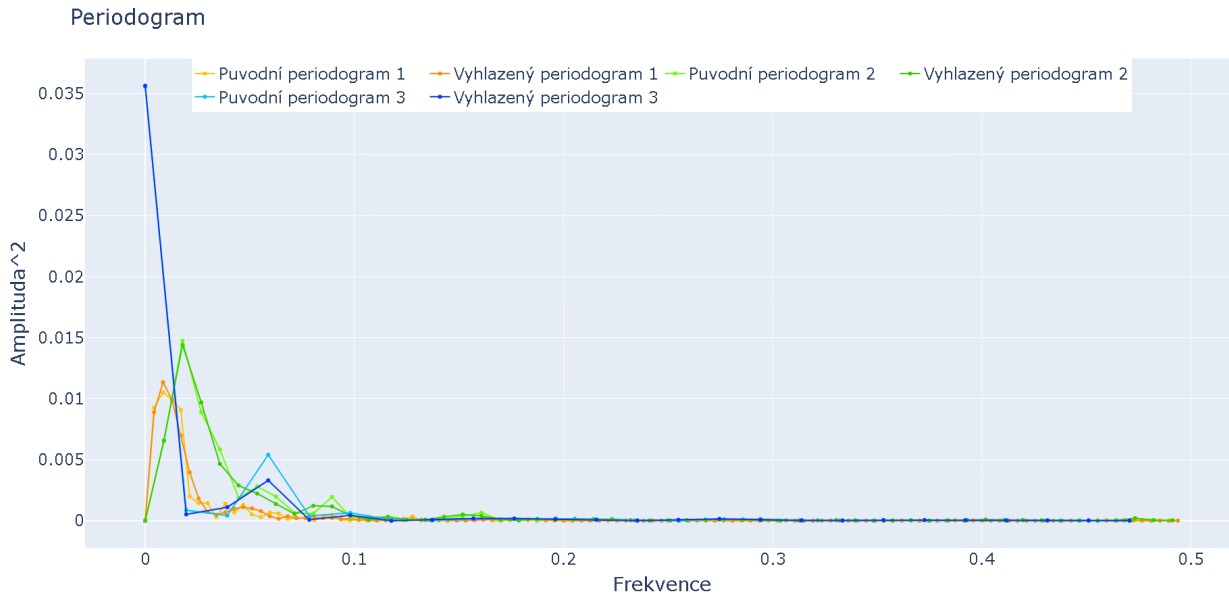
Obrázek 6.5: exponenciální vyrovnání Rolls Royce

6.2.3 Výpočet periodicity

Nyní tedy potřebujeme udělat spektrální analýzu na známých datech po odečtení trendů. Spektrální analýzu jsem dělal ve třech fázích, nejdříve jsem našel významné periody a vyrovnal data na celém časovém úseku, poté na druhé polovině vyrovnaných dat a nakonec na poslední čtvrtině dat. Důvodem k tomuto kroku byl fakt, že se některé periodické prvky mohou objevit až v průběhu roku a spektrální analýza přes celý časový interval takovou periodicitu neodhalí, na druhou stranu ale, pokud by nějaká periodicitu v průběhu roku zmizela, tak se do výsledného periodogramu již nedostane, což nám při predikci nevádí. Periodogramy jednotlivých fází jsem vykreslil do jednoho periodogramu, zde si musíme uvědomit, že vypočtené frekvence nejsou počítány na stejně dlouhých intervalech, tudíž stejná frekvence dvou různých křivek v tomto periodogramu (nepočítáme-li původní a vyrovnaný periodogram) neodpovídá stejné délce periody naší hledané periodické funkce.

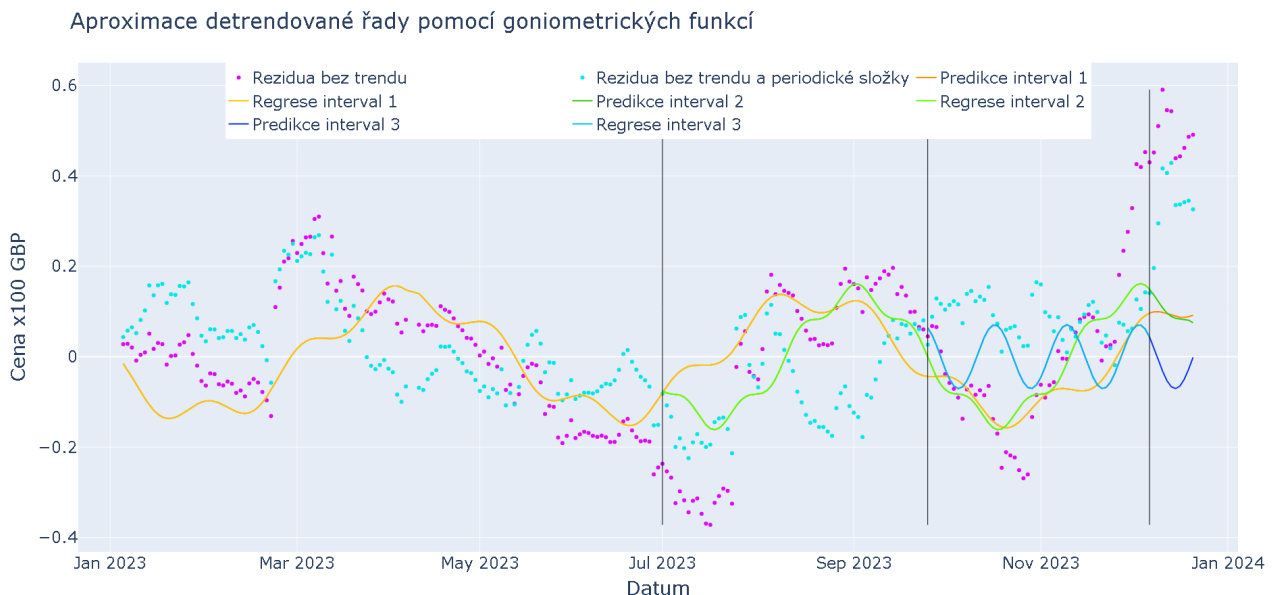
Začneme periodogramem pro lineární regresi 6.6. Jak jsem se již zmínil výše, na periodogramech si můžeme povšimnout, že kromě původních periodogramů, vykreslených světlejšími barvami, jsou zde periodogramy v tmavších odstínech, tyto periodogramy jsou vyhlazeny pomocí Savitzky-Golay filtru s řádem polynomu 9 a délkou okna 25, viz [12], tyto hodnoty jsem volil experimentálně tak, abych došel k nějakému rozumnému vyhlazení a ušetřil čas při samotné regresi periodických složek. Ze stejného důvodu jsem odfiltroval 5% nejméně významných period. Jak jsem zmínil v poznámce v teoretické části 5.1.3, filtrování a vyhlazování periodogramů je s ohledem na výpočetní náročnost regrese periodických složek velmi významným, ale složitým tématem, kterým jsem se pro jeho náročnost podrobně nezabýval.

Nakonec si můžeme všimnout, že u některých periodogramů se našla nenulová amplituda i u nulové frekvence, toto je způsobeno implementací použitého balíčku a znamená to, že hodnoty po odečtení trendu mají stále nenulovou střední hodnotu, do výpočtu periodických složek tuto frekvenci nezahrnujeme.



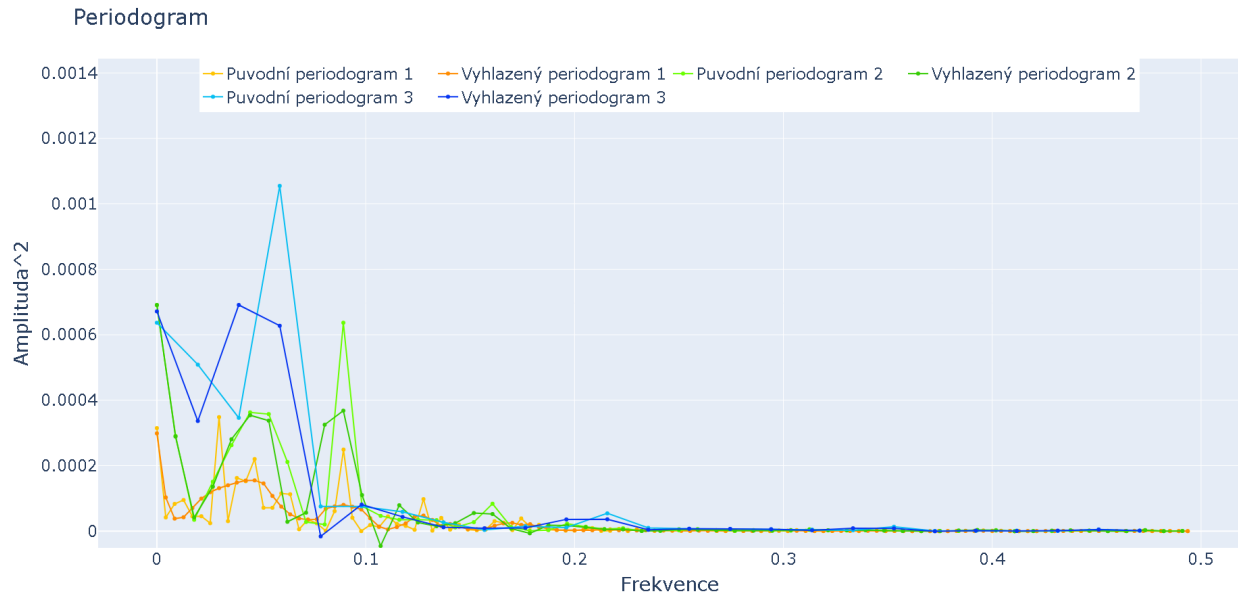
Obrázek 6.6: periodogram lineární regrese Rolls Royce

Nyní můžeme pomocí regrese a frekvencí získaných z periodogramu proložit naše detrendovaná data periodickými funkcemi pro každou fázi spektrální analýzy zvlášť 6.7. V této fázi již vkládám grafy i s předpovědí o našich 10 dní dopředu. Hranice předpovědi je označena zleva poslední svislou černou čarou. První dvě svislé čáry zleva označují postupně začátek druhé a třetí fáze hledání period.



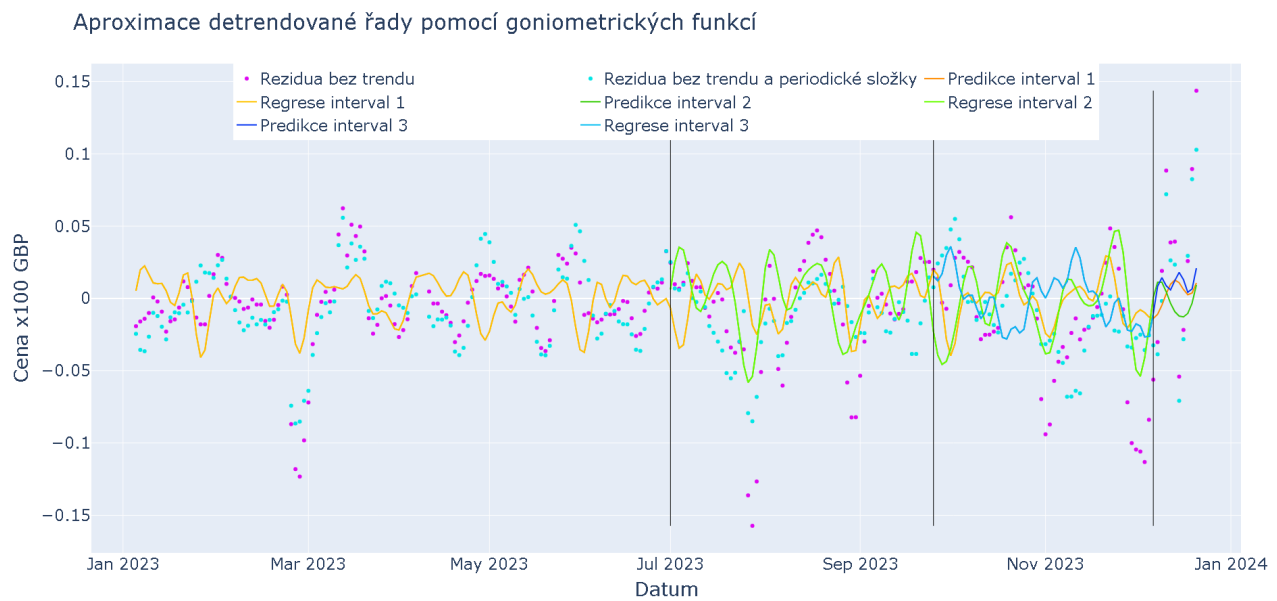
Obrázek 6.7: periodická složka lineární regrese Rolls Royce

Nyní spočítáme periodogramy také pro exponenciální vyrovnání, zde vkládám pouze periodogram čtyřnásobného vyrovnání 6.8, protože vím, že v tomto případě vychází čtyřnásobné vyrovnání nejlépe.



Obrázek 6.8: periodogram čtyřnásobného EMA Rolls Royce

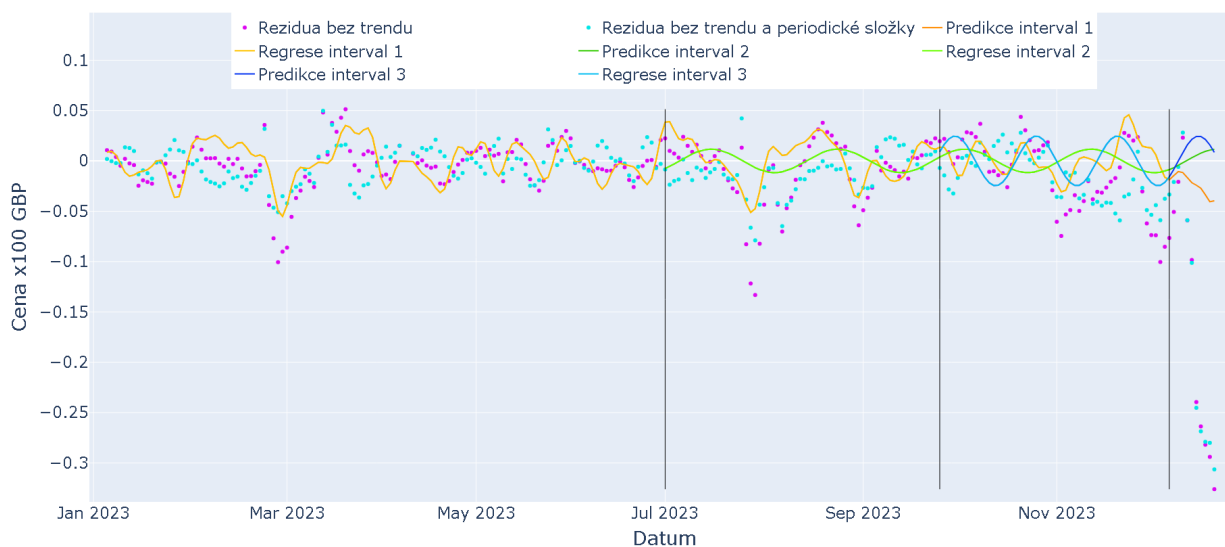
Na dalších čtyřech obrázcích poté můžeme vidět srovnání periodických složek. Nejdříve periodické složky čtyřnásobného vyrovnání 6.9.



Obrázek 6.9: periodická složka čtyřnásobného EMA Rolls Royce

Poté periodická složka trojnásobného vyrovnání 6.10. Zde vidíme, že periodicitu zjištěnou ve druhé fázi spektrální analýzy je oproti dalším dvěma složkám poměrně nevýznamná.

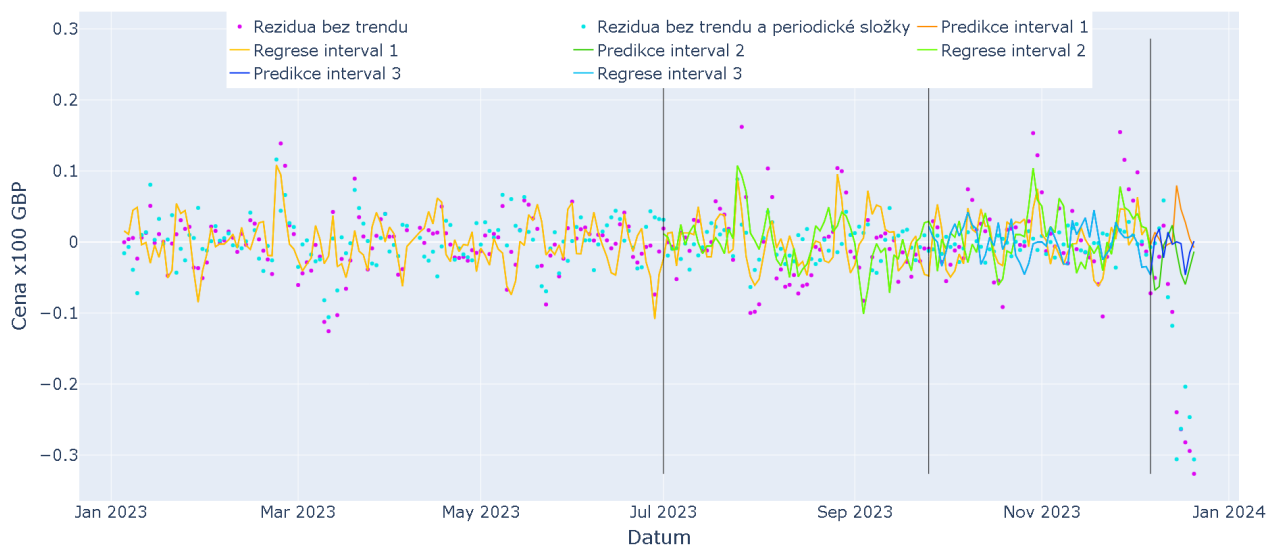
Aproximace detrendované řady pomocí goniometrických funkcí



Obrázek 6.10: periodická složka trojnásobného EMA Rolls Royce

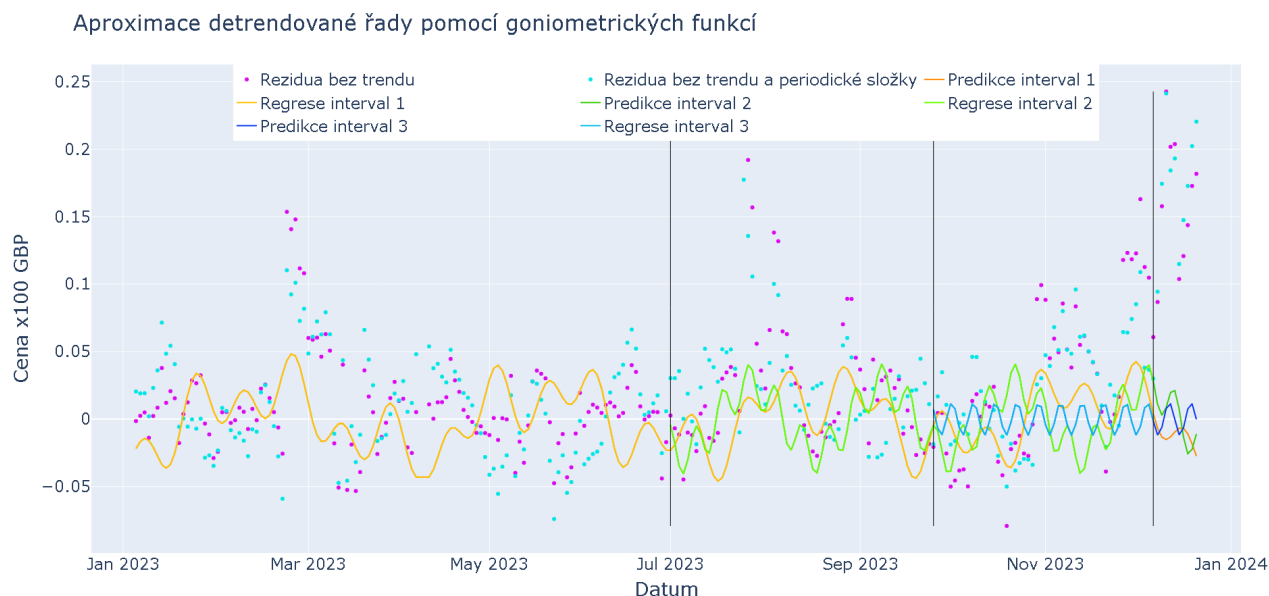
Dále zde máme periodickou složku dvojnásobného vyrovnání 6.11.

Aproximace detrendované řady pomocí goniometrických funkcí



Obrázek 6.11: periodická složka dvojnásobného EMA Rolls Royce

Nakonec periodická složka jednoduchého vyrovnání 6.12. Zde si můžeme všimnout, že periodičita nalezená v poslední fázi má nejmenší amplitudu, tudíž i nejmenší význam, což bychom mohli očekávat, jelikož jsme v předchozích dvou fázích již většinu periodity odstranili.

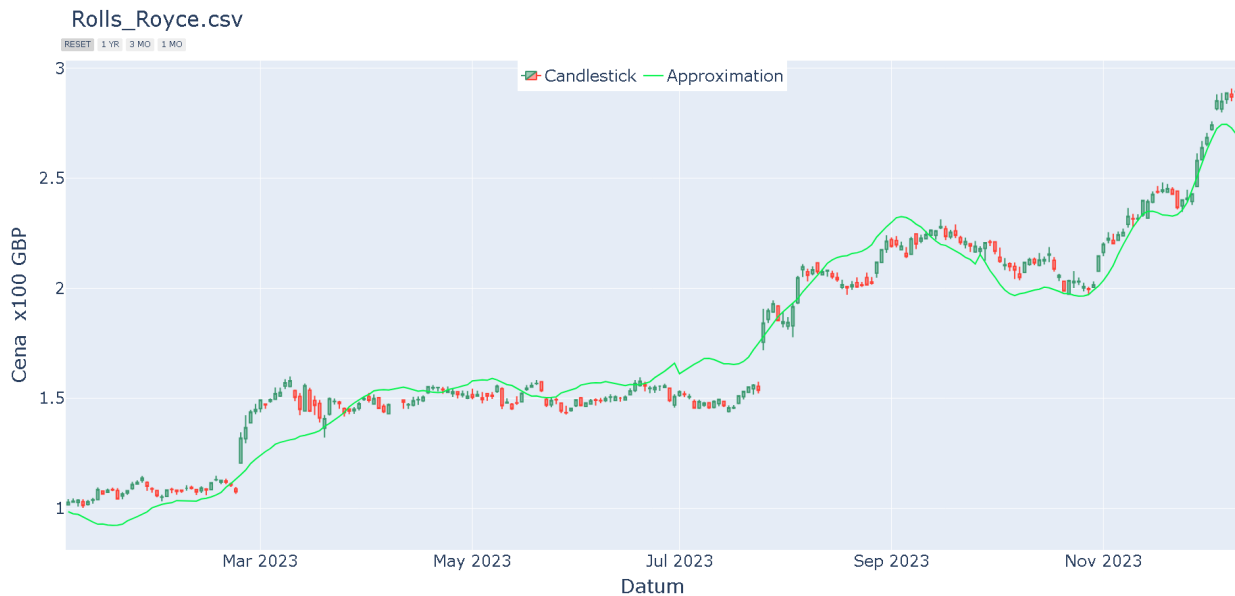


Obrázek 6.12: periodická složka jednoduchého EMA Rolls Royce

Zde si můžeme především povšimnout poměrně velkého rozdílu mezi amplitudami periodických složek těchto exponenciálních a vyrovnání a amplitudami periodické složky po lineární regresi 6.7, což je dle našeho očekávání. Další patrnou věcí z těchto grafů je, že reziduální složky jdou i po vyrovnání výrazně dál od nulové střední hodnoty u tří exponenciálních vyrovnání nižších řádů, než u řádu čtvrtého. Zde, ale musíme mít stále na paměti, že tento vývoj nedokážeme spolehlivě předpovědět.

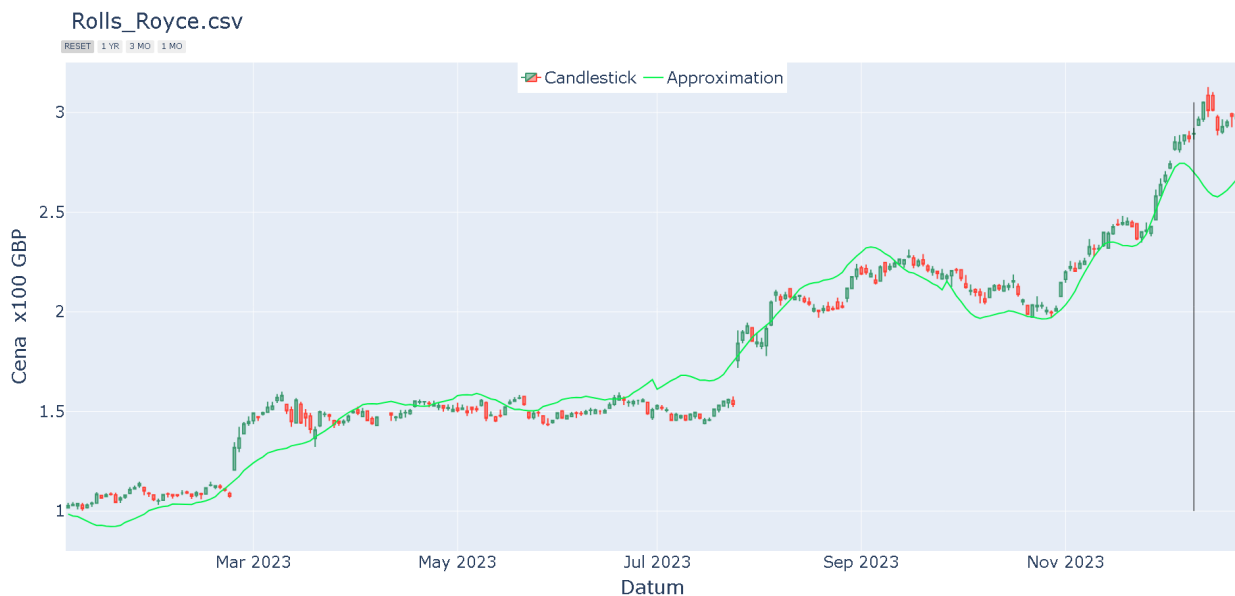
6.2.4 Kompletní model

Nakonec modely trendové složky složíme dohromady s modely periodických složek a podíváme se na výsledky. Nejdříve zde máme graf kompletního modelu za použití lineární regrese bez předpovědi 6.13.



Obrázek 6.13: lineární regrese s periodicitou Rolls Royce před předpovědí

Dále lineární model s předpovědí 6.14 (případně interaktivní graf 8).



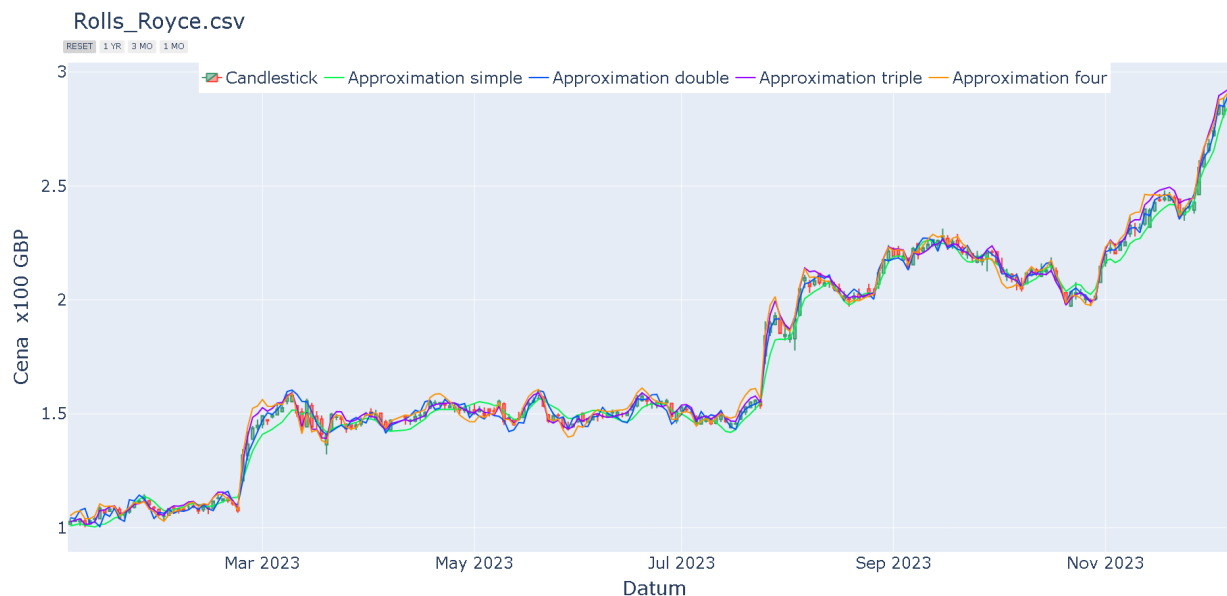
Obrázek 6.14: lineární regrese s periodicitou Rolls Royce s předpovědí

Nakonec detail tohoto modelu na vývoj v posledních třech měsících 6.15. Z grafů s předpovědí je patrné, že graf s lineární regresí očekával vlivem periodické složky pokles, zatímco ve skutečnosti nastal mírný růst, nebo spíše pohyb doprava (tj. pohyb, kdy časová řada osciluje kolem konstantního trendu na daném časovém úseku).



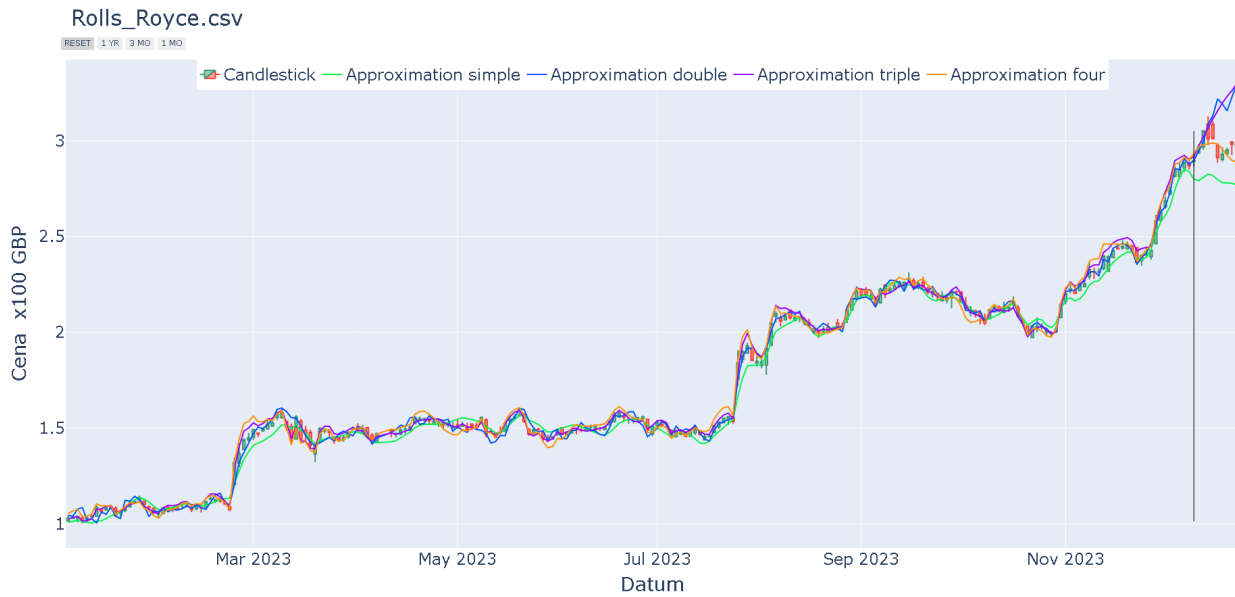
Obrázek 6.15: lineární regrese s periodicitou Rolls Royce s předpovědí poslední 3 měsíce

Nyní se můžeme podívat na porovnání jednotlivých řádů exponenciálních vyrovnání, které jsem pro lepší porovnání vložil do jednoho grafu. Nejdříve na 6.30 vidíme regresi na známých datech.



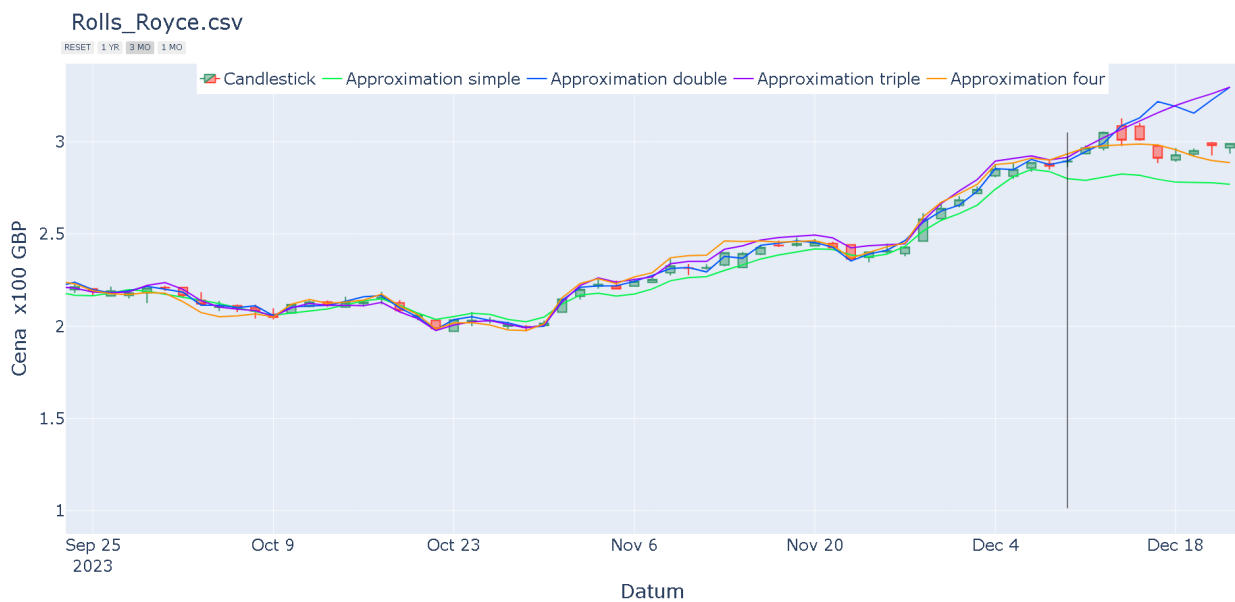
Obrázek 6.16: exponenciální vyrovnání s periodicitou Rolls Royce před předpovědí

A nakonec na grafu celé řady 6.17 (případně interaktivním grafu 7)



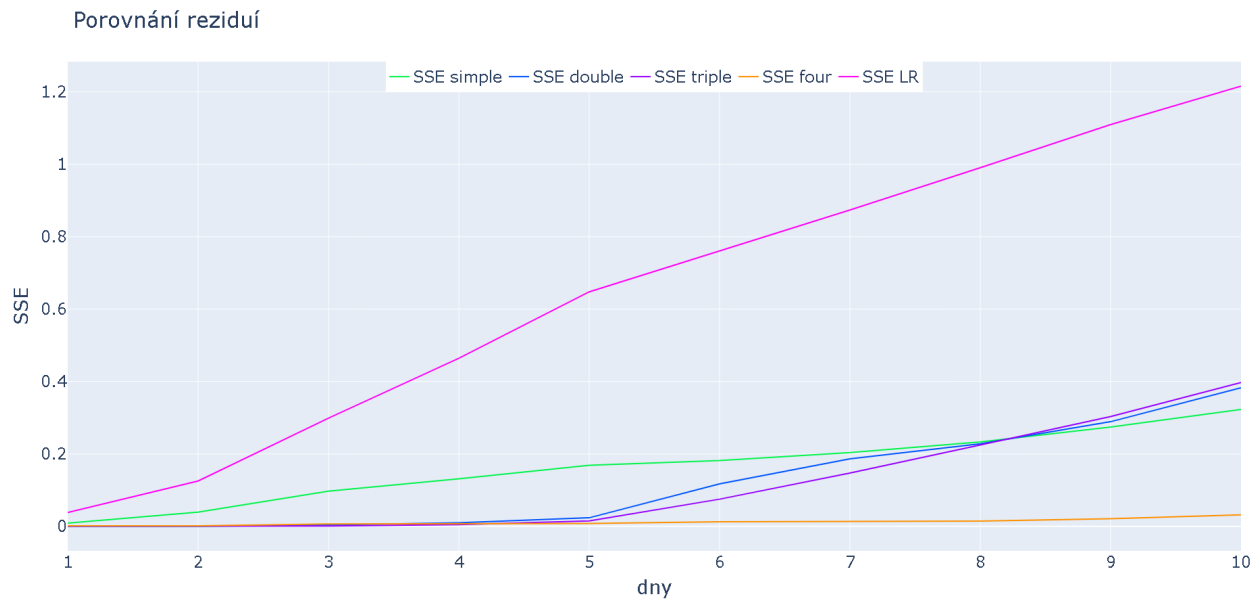
Obrázek 6.17: exponenciální vyrovnání s periodicitou Rolls Royce s předpovědí

a pak především na grafu posledních tří měsíců 6.18 můžeme vidět naši předpověď.



Obrázek 6.18: exponenciální vyrovnání s periodicitou Rolls Royce s předpovědí poslední 3 měsíce

Jak vidíme z grafů s předpověďmi a také z porovnání sum reziduálních čtverců na 6.19, v tomto konkrétním případě bylo nejlepší volbou čtyřnásobné exponenciální vyrovnání



Obrázek 6.19: porovnání kvadrátů chyb Rolls Royce

6.3 Akcie Admiral Group

6.3.1 Výchozí data

Dalším grafem, který zde uvedu jsou akcie finanční společnosti Admiral Group plc. Opět nejdříve graf známých dat 6.20.



Obrázek 6.20: akcie Admiral group 1 rok

A následně časová řada i s hodnotami, které bychom měli předpovědět 6.21.

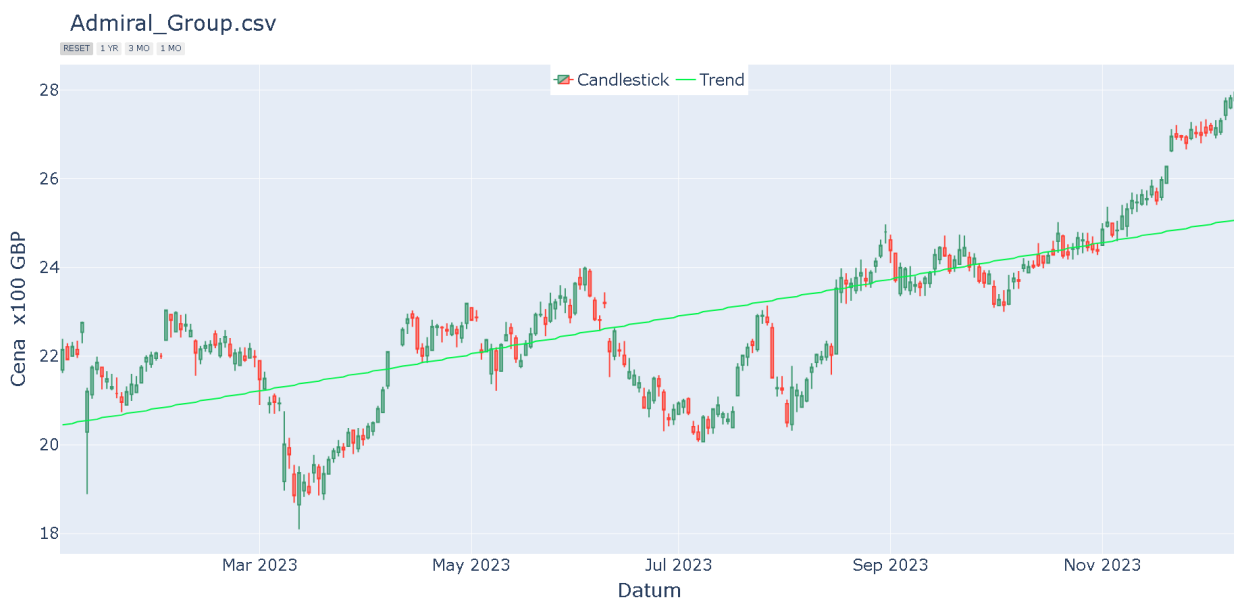


Obrázek 6.21: akcie Admiral group 1 rok s předpovídanými hodnotami

Zatímco v případě akcií Rolls-Royce 6.2 jsme mohli v predikovaných datech vidět zvlnění a pohyb doprava, nyní máme v předpovídaných datech 6.21 obrat trendu a pohyb dolů.

6.3.2 Výpočet trendu

Časovou řadu proložíme nejdříve lineární regresní přímkou 6.22.



Obrázek 6.22: lineární regrese Admiral group bez předpovědi

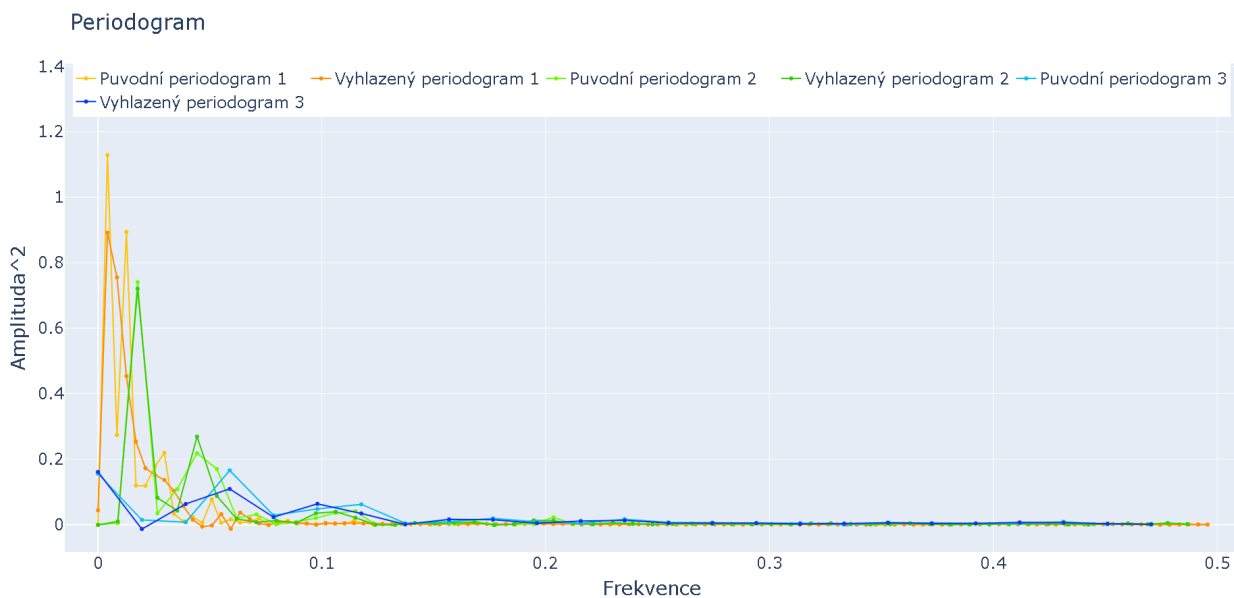
A poté uděláme exponenciální vyrovnání prvního až čtvrtého řádu 6.23.



Obrázek 6.23: exponenciální vyrovnání Admiral group bez předpovědi

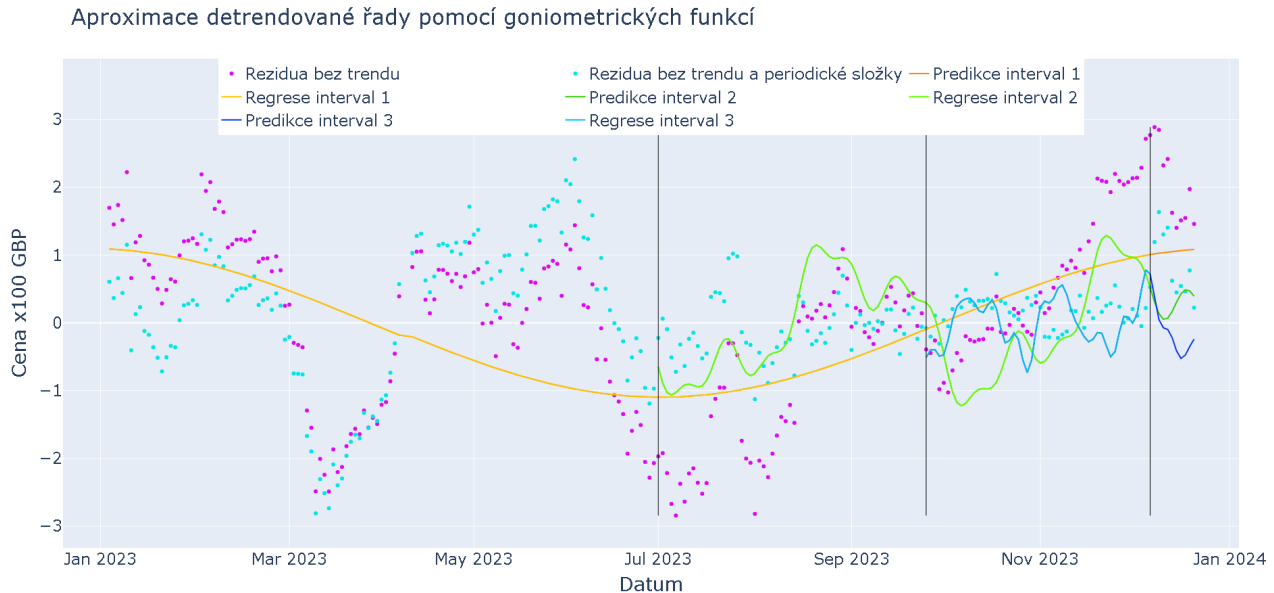
6.3.3 Výpočet periodicity

Zde vyjdu opět z výsledného kompletního modelu a ukáži zde pouze periodicity dvou modelů. Nejdříve lineárního modelu 6.24, kde po vyhlazení zůstane v první fázi pouze nejdelší možná periodická složka a zároveň v druhé fázi vznikne periodická složka s podobně velkou amplitudou, což poté ve výsledném grafu způsobí zajímavý zlom 6.28.



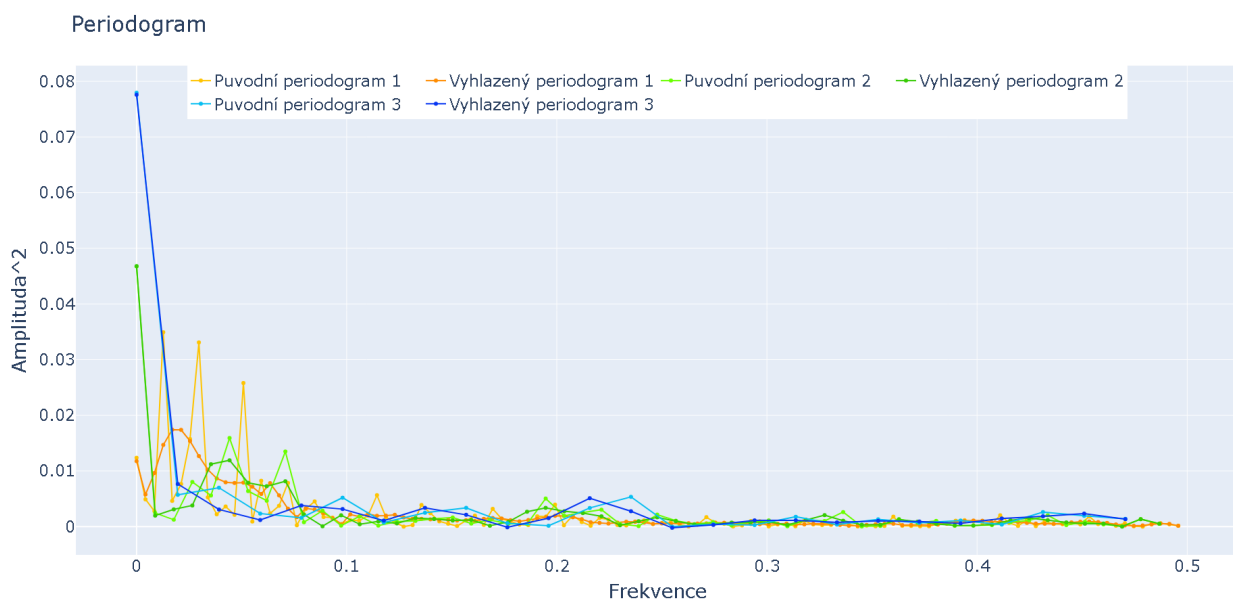
Obrázek 6.24: periodogram lineární regrese Admiral group

Zde máme naše periodické složky lineární regrese v jednom grafu 6.25. Na tomto grafu si můžeme všimnout významu naší třífázové spektrální analýzy, všechny tři periodické složky jsou téměř stejně významné, ale v první fázi se zjistila pouze malá část celkové periodicity.



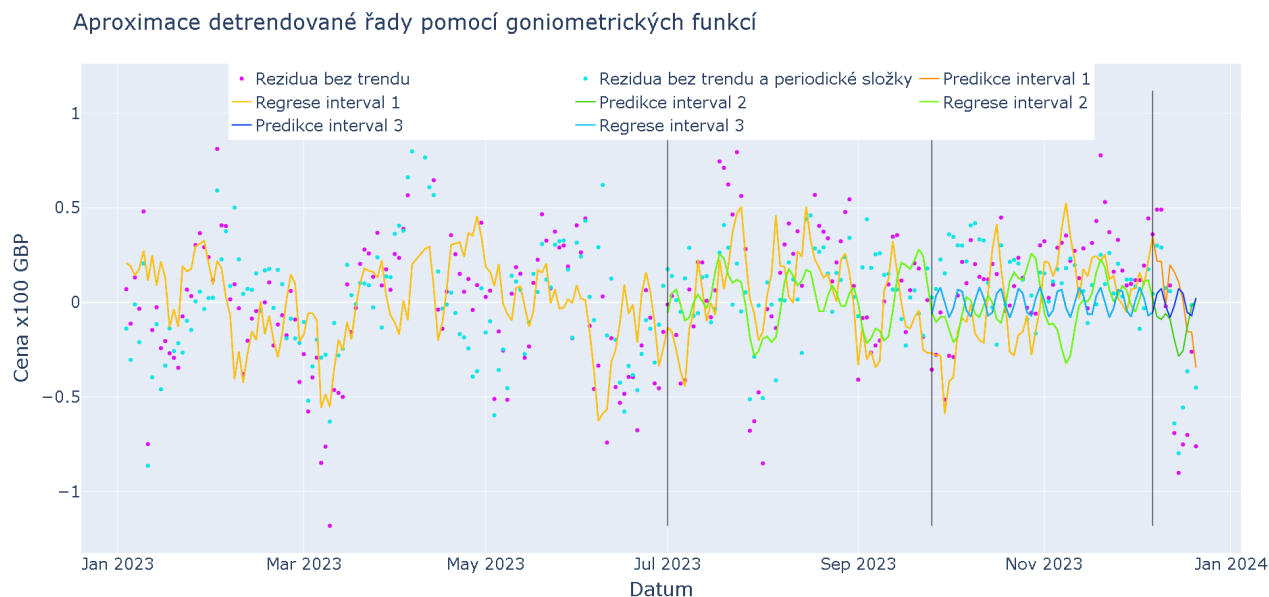
Obrázek 6.25: periodická složka lineární regrese Admiral group

Následně vykreslíme periodogram po jednoduchém exponenciálním vyrovnaní 6.26.



Obrázek 6.26: periodogram jednoduchého EMA Admiral group

A výsledné periodické složky 6.27 jednoduchého exponenciálního vyrovnaní.



Obrázek 6.27: periodická složka jednoduchého EMA Admiral group

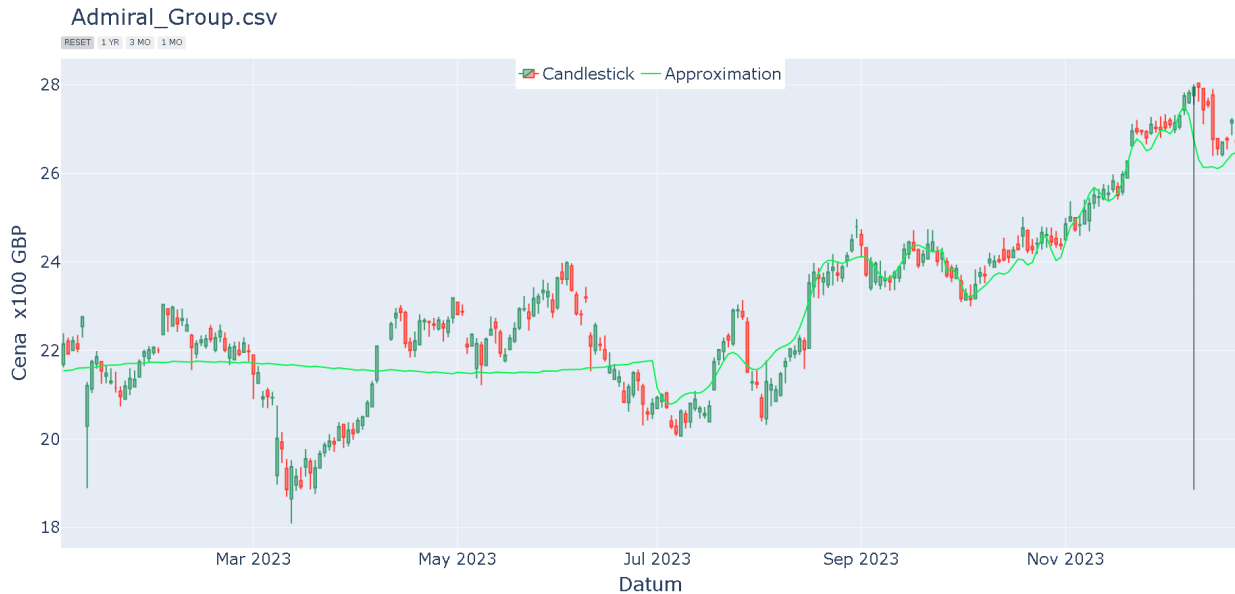
6.3.4 Kompletní model

Zde můžeme v kompletním modelu lineární regrese s periodickou složkou 6.28 vidět zmíněný zlom v grafu způsobený přidáním periodicity z druhé fáze spektrální analýzy. Zároveň si můžeme povšimnout, že zatímco v první polovině grafu, kde je pouze periodicitu z první fáze, je vyrovnaná křivka od skutečných dat poměrně dost vzdálená, tak ve druhé a zejména ve třetí fázi již poměrně dobře kopíruje skutečná data.



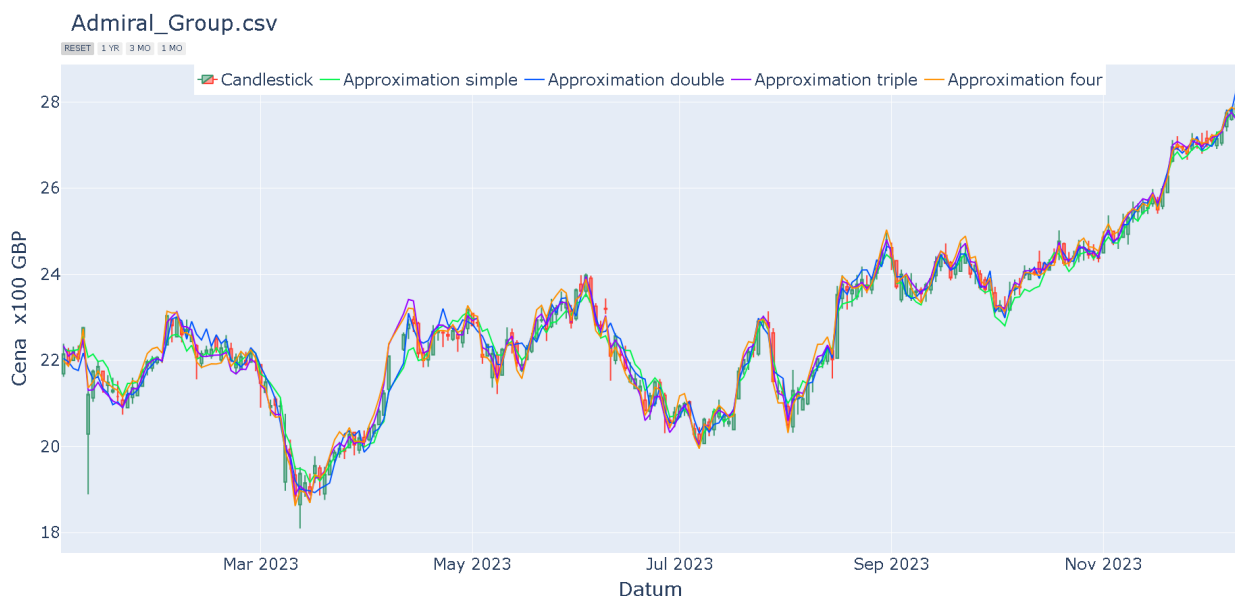
Obrázek 6.28: lineární regrese s periodicitou Admiral Group před předpovědí

Na dalším obrázku můžeme vidět data již s předpovědí 6.29 (případně interaktivní graf 2), ze kterých je patrné, že naše předpověď na základě regrese správně předpověděla pokles, ale nadhodnotila jeho rozsah.



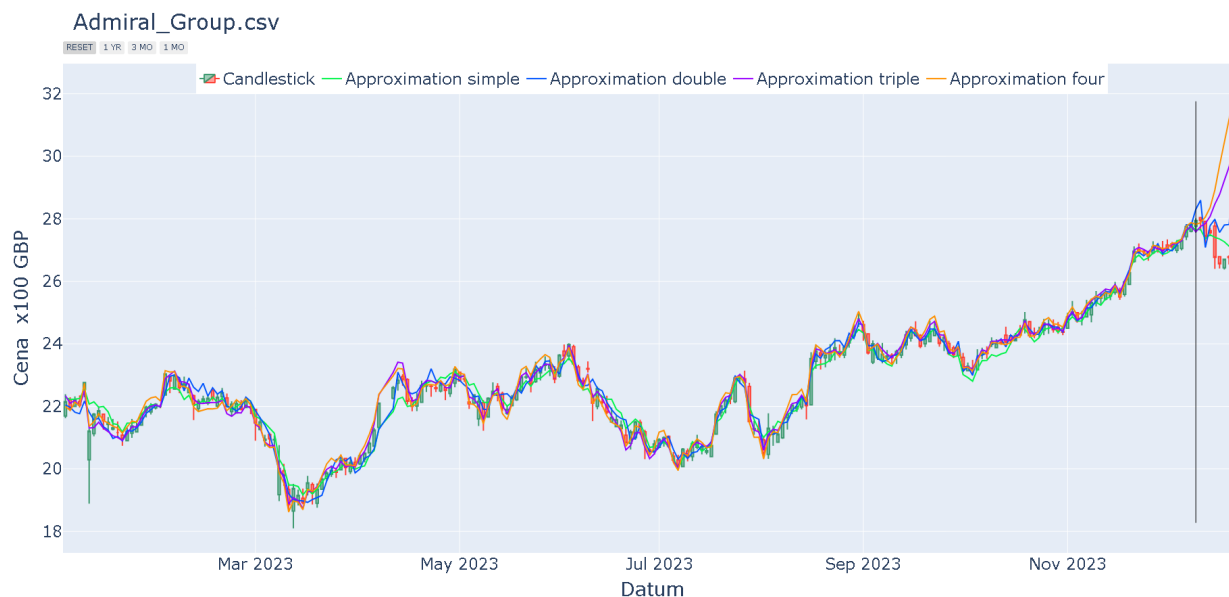
Obrázek 6.29: lineární regrese s periodicitou Admiral Group s předpovědí

Nyní se přesuneme k exponenciálnímu vyrovnání 6.30.



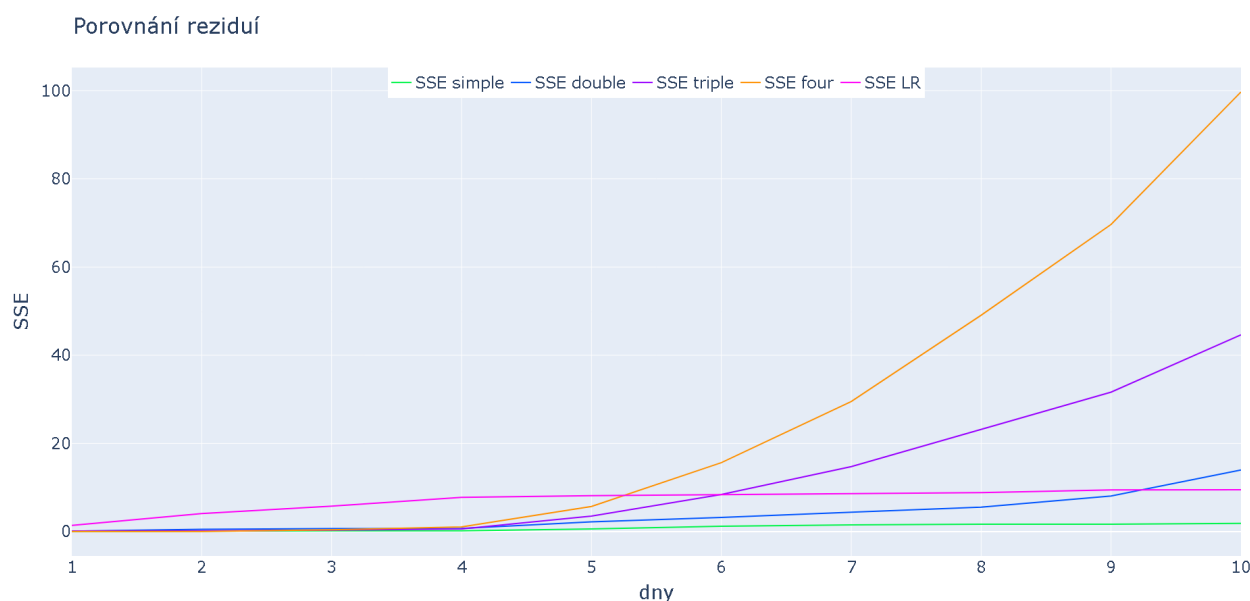
Obrázek 6.30: exponenciální vyrovnání s periodicitou Admiral Group před předpovědí

Zde 6.31 (případně v interaktivním grafu 1) je z předpovědí patrné, že exponenciální vyrovnání vyššího řádu, než řádu jedna, tj. jednoduchého exponenciálního vyrovnání, které je konstantní v předpovědi, nemohly zlom, který nastal až v budoucnosti, předpovědět, to je jejich charakteristickou vlastností. Důvodem, proč zde fungovalo jednoduché vyrovnání nejlépe je to, že jeho periodická složka ho dokázala dostatečně nasměrovat dolů, ale nebyla tak výrazná jako u lineární regrese.



Obrázek 6.31: exponenciální vyrovnání s periodicitou Admiral Group s předpovědí

V tomto případě výsledek 6.32 významně ovlivnily první čtyři předpovídané dny, kdy lineární regrese odhadovala pokles dříve a významnější, než nastal. Dále také v případě jednoduchého exponenciálního vyrovnání způsobily periodické složky dostatečně silný pohyb správným směrem, díky čemuž na tomto časovém úseku vyšlo jednoduché exponenciální vyrovnání nejlépe.



Obrázek 6.32: porovnání kvadrátů chyb Admiral group

6.4 Akcie BAE systems

6.4.1 Výchozí data

Dostáváme se ke třetímu ukázkovému grafu 6.33.



Obrázek 6.33: akcie BAE 1 rok

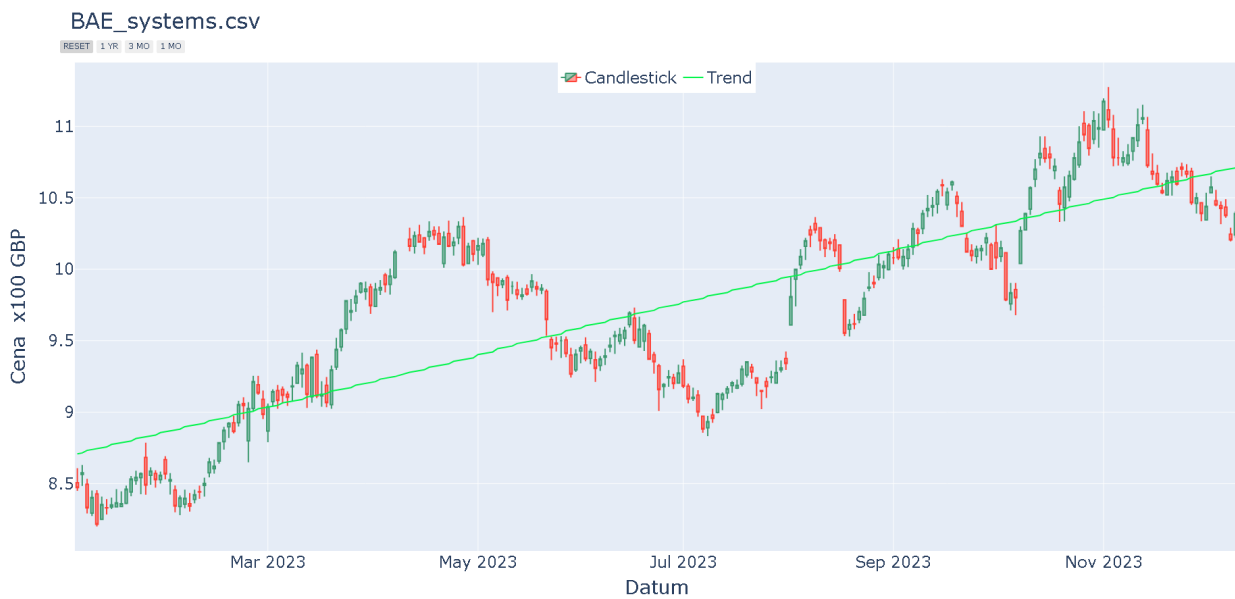
V grafu 6.34 vidíme, že nastává zlom právě v první předpovídané hodnotě, to znamená, že exponenciální vyrovnání vyšší řádů zde pravděpodobně nebudou fungovat, protože z předchozích dat bude vyplývat pokles.



Obrázek 6.34: akcie BAE 1 rok s předpovídanými hodnotami

6.4.2 Výpočet trendu

Vypočteme lineární regresi 6.35.



Obrázek 6.35: lineární regrese BAE bez předpovědi

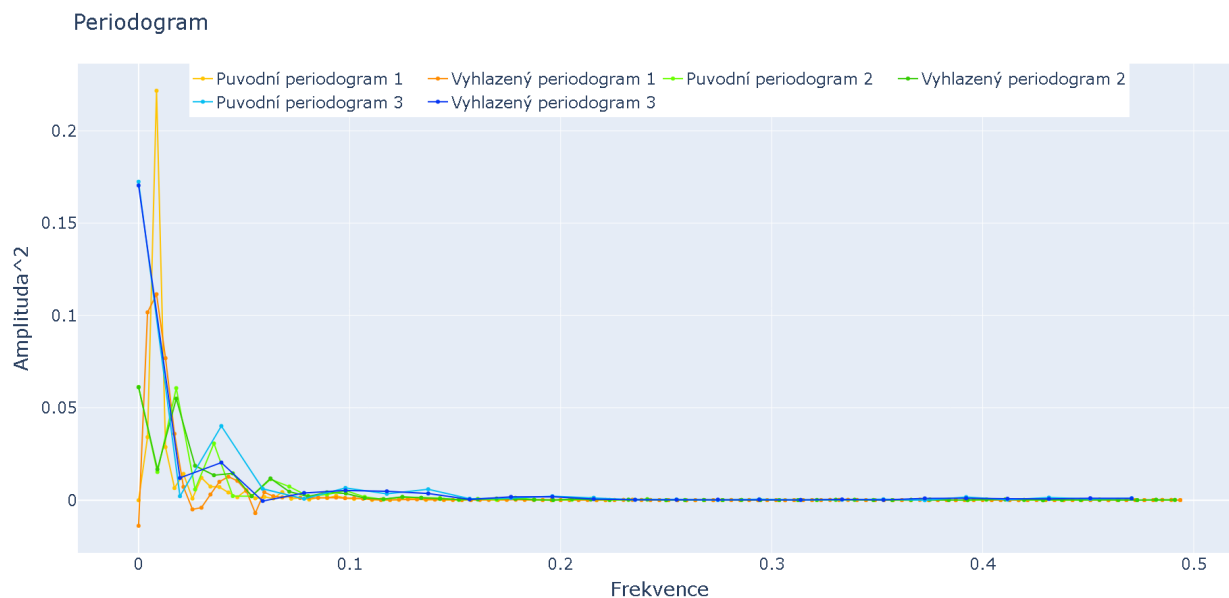
A poté exponenciální vyrovnání 6.36.



Obrázek 6.36: exponenciální vyrovnání BAE bez předpovědi

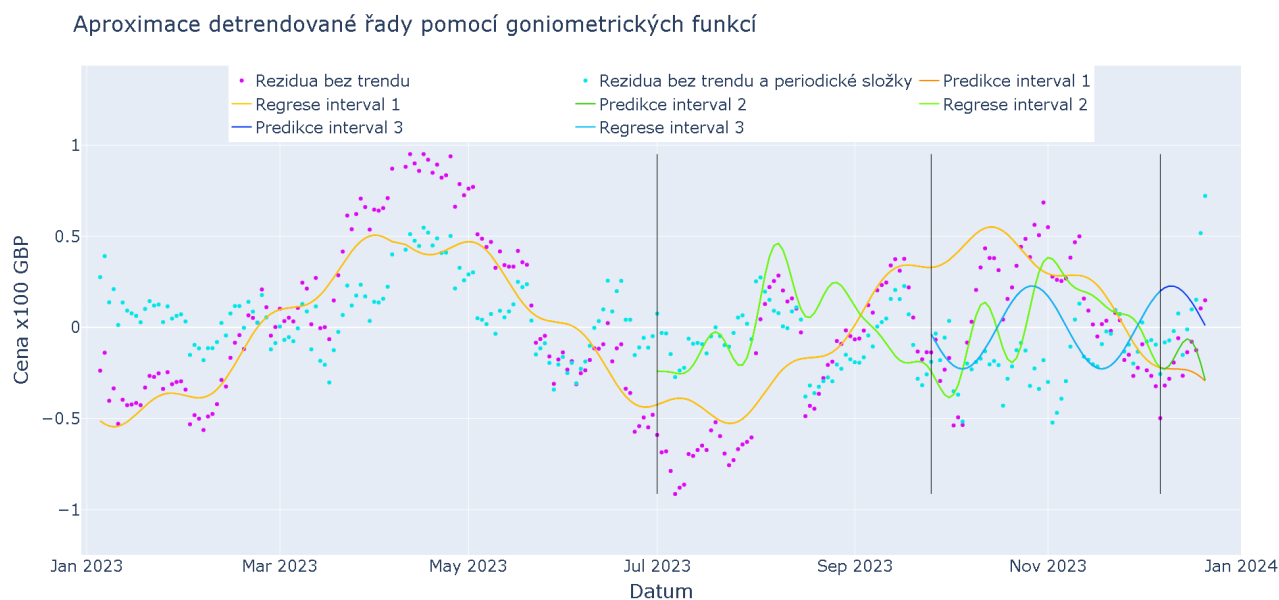
6.4.3 Výpočet periodicity

Je zřejmé, že zde může vycházet nejlépe buď lineární regrese, nebo jednoduché exponenciální vyrovnání. Nejdříve spočteme periodogram časové řady po lineární regresi 6.37.



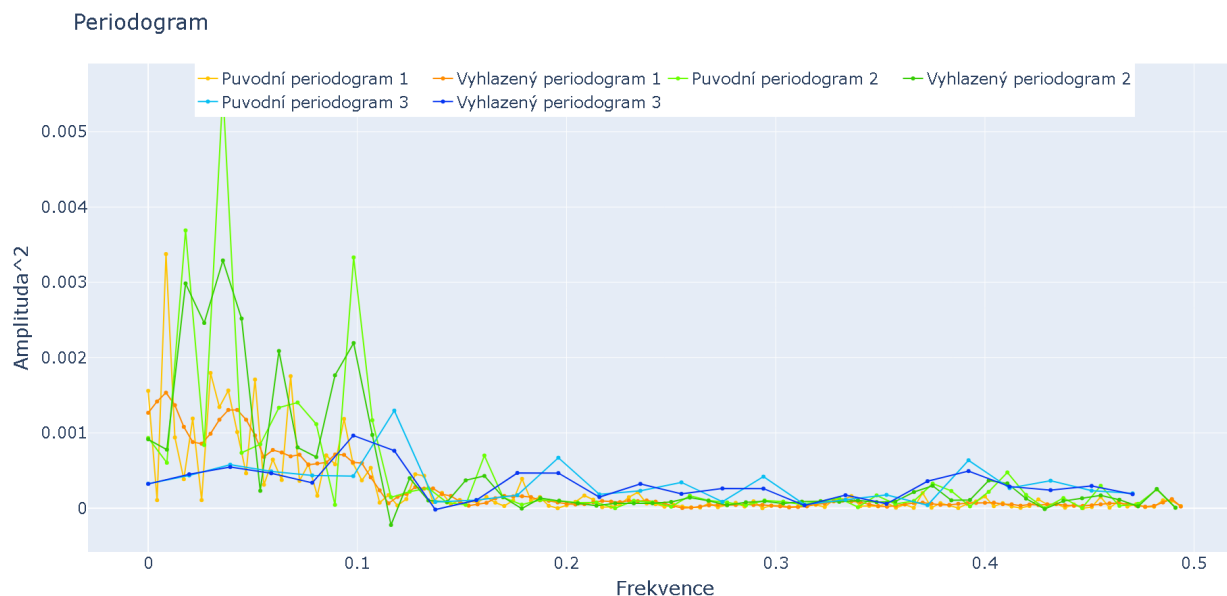
Obrázek 6.37: periodogram lineární regrese BAE

Následně pomocí dat z periodogramu najdeme periodické složky 6.38.



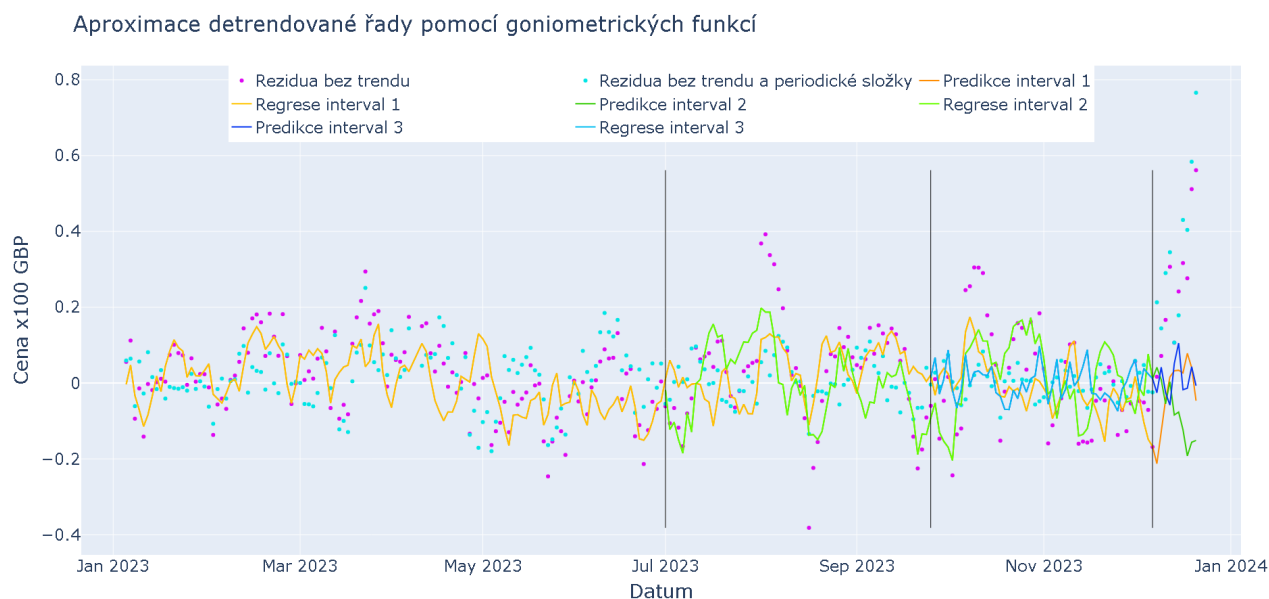
Obrázek 6.38: periodická složka lineární regrese BAE

Stejně postupujeme u jednoduchého exponenciálního vyrovnání, nejdřív periodogram 6.39



Obrázek 6.39: periodogram lineární regrese BAE

A poté periodické složky 6.40.

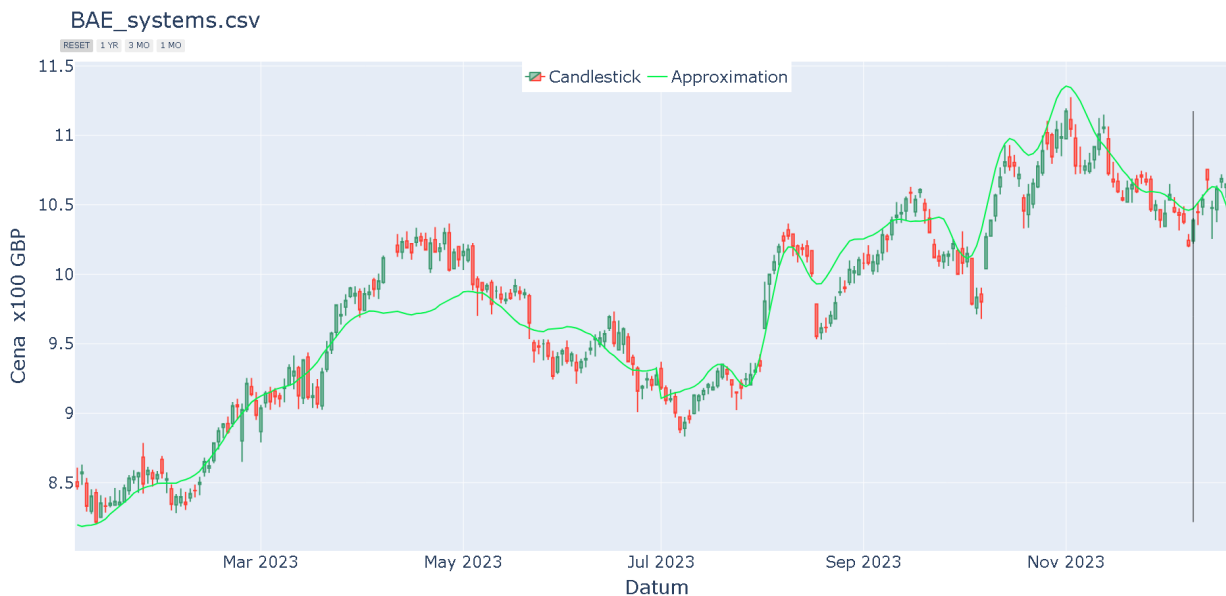


Obrázek 6.40: periodická složka lineární regrese BAE

Při porovnání periodických složek po lineární regresi 6.38 a periodických složek jednoduchého exponenciálního vyrovnání 6.40 vidíme, že periodická složka bude mít značně větší vliv v případě lineární regrese.

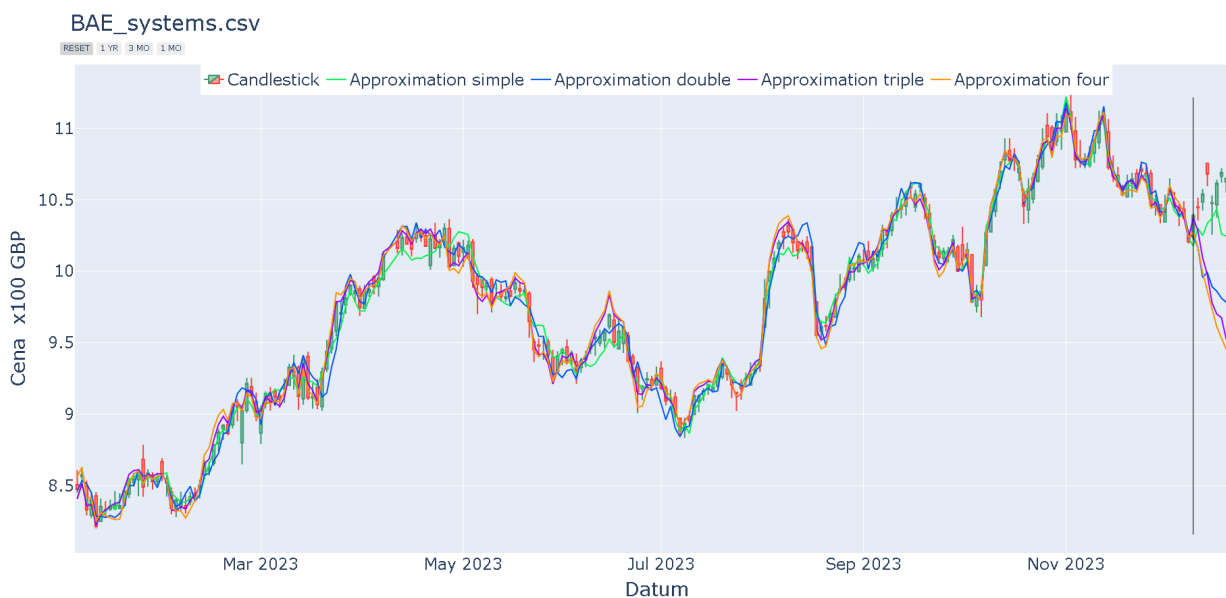
6.4.4 Kompletní model

V kompletním modelu lineární regrese s periodicitou 6.41 (případně v interaktivním grafu 4) vidíme, že periodicitu tento model nejdříve přiblížila a poté se zlomila opačným směrem.



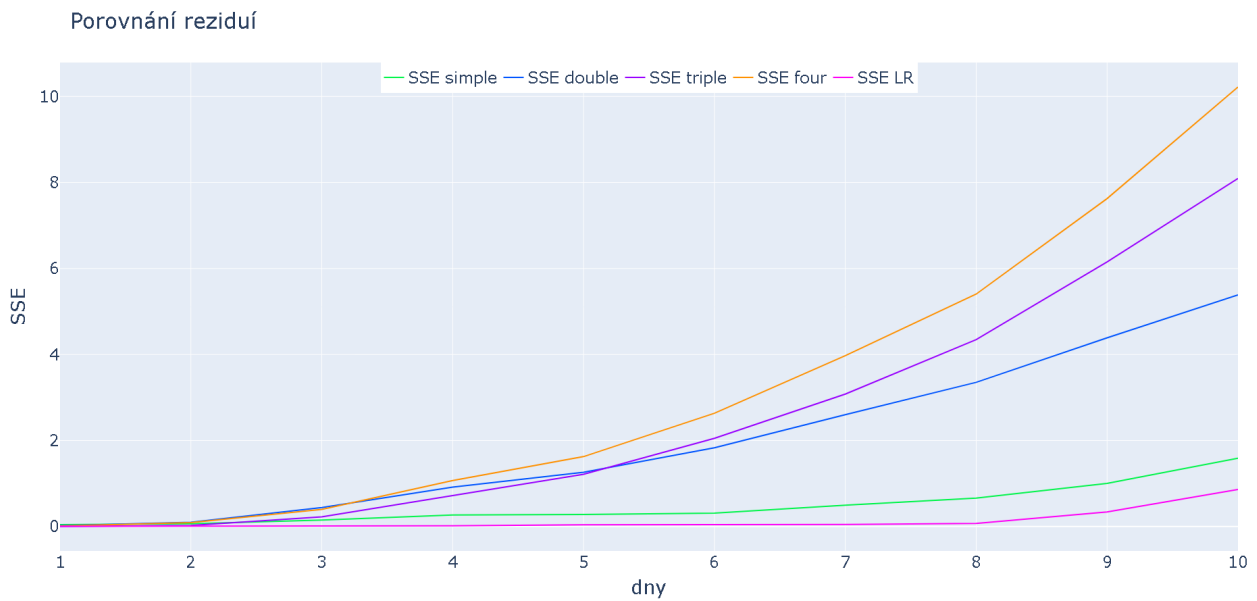
Obrázek 6.41: lineární regrese s periodicitou BAE s předpovědí

V případě jednoduchého exponenciálního vyrovnání 6.42 (interaktivní graf 3) došlo k tomuto jevu také, ale v menší míře.



Obrázek 6.42: exponenciální vyrovnání s periodicitou BAE s předpovědí

Z výsledného porovnání chyb 6.43 vidíme, že nejlepších výsledků dosáhla lineární regrese. Nicméně i tento model dokázal poměrně uspokojivě předpovědět pouze prvních 6 kroků a poté se jeho vývoj vlivem periodických složek obrátil opačným směrem.



Obrázek 6.43: porovnání kvadrátů chyb BAE

7 Závěr

V této práci jsem se zabýval studiem a testováním některých způsobů dekompozice časových finančních řad. Do kapitoly přílohy jsem vložil interaktivní grafy některých časových řad, kterými jsem se zabýval v podobě html odkazů pro lepší přehled.

Nejdříve jsem v druhé kapitole uvedl nějaké základní pojmy důležité při studiu finančních časových řad, jako jsou některé statistické pojmy, a zejména pak mnou používané druhy grafů. V druhé kapitole jsem potom rozebral typy technické analýzy jako stručný přehled možných přístupů k tomuto úkolu. Dále čtvrtou kapitolu jsem věnoval mnou použitým způsobům odstranění trendové složky z časové řady, konkrétně lineární regresi a exponenciálním vyrovnáním. Obsahem páté kapitoly byly stručné matematické základy spektrální analýzy. A nakonec v kapitole číslo šest jsem tyto znalosti uplatnil na reálné finanční časové řady a ukázal na některých těchto řadách využití daného aparátu nejdříve pro analýzu známých dat a především pak pro možnosti predikce, kde jsem okomentoval vlastnosti, výhody a nevýhody těchto přístupů.

Zde musím zmínit, že výsledky prezentované v kapitole šest jsou pouhým zlomkem výsledků, které jsem získal, ale byly vybrány tak, aby co nejlépe prezentovaly celek.

Mým závěrem je, že finanční trhy nelze tímto způsobem efektivně analyzovat a následně predikovat. Jak jsem ukázal v praktické části, žádný z těchto způsobů není dostatečně efektivní a samotná spektrální analýza je potom výpočetně velmi náročná, kvůli regresi periodických složek, což nás nutí k používání různých druhů filtrů a samo o sobě to posléze snižuje efekt této analýzy. Nakonec finanční trhy jsou velmi komplexní obor a bylo by naivní si myslet, že jedna daná časová řada jde vysvětlit sama o sobě pouze za použití technické analýzy a bez dalších vlivů. Proto bych se v dalším studiu chtěl zabývat korelacemi mezi více časovými řadami najednou a dále tak rozvíjet a zdokonalovat tento model.

8 Přílohy

1. Admiral Group EMA diagram
2. Admiral Group LR diagram
3. BAE systems EMA diagram
4. BAE systems LR diagram
5. Barclays EMA diagram
6. Barclays LR diagram
7. Rolls Royce EMA diagram
8. Rolls Royce LR diagram
9. Tesco EMA diagram
10. Tesco LR diagram
11. Vodafone EMA diagram
12. Vodafone LR diagram

Literatura

- [1] Anděl, Jiří, *Základy matematické statistiky*, vyd. 3., Praha: Matfyzpress, 2011.
- [2] Box, George Edward Pelham, Jenkins, Gwilym Meirion, *Time series analysis, forecasting and control*, 1.vyd., San Francisco, Holden Day, 1970.
- [3] Brown, Robert Goodell, *Smoothing, forecasting and prediction of discrete time series*, Prentice-Hall, Inc., 1963.
- [4] Cipra, Tomáš, *Analýza časových řad s aplikacemi v ekonomii*, 1. vyd., Praha: SNTL - Nakladatelství technické literatury, 1986.
- [5] Cooley, James William, Lewis, A. W. Peter, Welch, D. Peter, *The fast Fourier transform algorithm and its applications.*, Res. Monograph RC 1743, IBM Watson Res., Center, New York, 1967.
- [6] Cooley, James William, Tukey, John Wilder, *An algorithm for the machine computation of complex Fourier series.*, Math Comput 19:297–301, 1965.
- [7] Dhrymes, Phoebus James, *Econometrics*, Springer, New York, 1974.
- [8] Reitano, R. Robert, *Foundations of Quantitative Finance: 3. The Integrals of Lebesgue and (Riemann-)Stieltjes*, Brandeis International Business School, Waltham, 2017.
- [9] Knihovna Numpy. online. Dostupné z: <https://numpy.org/doc/stable/> [citováno 2024-05-19].
- [10] Knihovna Pandas. online. Dostupné z: <https://pandas.pydata.org/docs/> [citováno 2024-05-19].
- [11] Knihovna Plotly. online. Dostupné z: <https://plotly.com/python/> [citováno 2024-05-19].
- [12] Savitzky, Abraham, Golay, Marcel J.E., "Smoothing and Differentiation of Data by Simplified Least Squares Procedures,"Anal. Chem. 1964, str. 1627-1639.
- [13] Knihovna Scipy. online. Dostupné z: <https://docs.scipy.org/doc/scipy/> [citováno 2024-05-19].
- [14] Schuster, A., *On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. Terrestrial magnetism*, III, 1898.
- [15] Shumway, R. H., Stoffer, D. S. *Time series analysis and its applications: with R examples*, 4th ed., Springer International Publishing, 2017.
- [16] Pilgrim, Mark, Willison, Simon *Dive Into Python 3*, vol. 2, Springer, 2009