

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA

**BAKALÁŘSKÁ PRÁCE**

Možnosti predikce ve finančních časových řadách



**Katedra matematické analýzy a aplikací matematiky**

Vedoucí bakalářské práce: **Mgr. Ondřej Vencálek, Ph.D.**

Vypracoval(a): **Zuzana Mikolašová**

Studijní program: B1103 Aplikovaná matematika

Studijní obor Matematika–ekonomie se zaměřením na bankovníctví/pojišťovnictví

Forma studia: prezenční

Rok odevzdání: 2018

## BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Zuzana Mikolašová

**Název práce:** Možnosti predikce ve finančních časových řadách

**Typ práce:** Bakalářská práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** Mgr. Ondřej Vencálek, Ph.D.

**Rok obhajoby práce:** 2018

**Abstrakt:** Tato bakalářská práce se bude zabývat statistickou analýzou finančních časových řad, jejichž hodnoty predikujeme pomocí statistických modelů. Při modelování využijí zejména dekompozici časových řad. Cílem této práce bude navržení vhodného modelu pro konkrétní finanční časovou řadu a zároveň porovnání predikcí s reálnými daty.

**Klíčová slova:** finanční časová řada, statistický model, dekompozice, predikce, martingal

**Počet stran:** 66

**Počet příloh:** 0

**Jazyk:** český

## BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Zuzana Mikolašová

**Title:** Cappability of predictions in financial time series

**Type of thesis:** Bachelor's

**Department:** Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** Mgr. Ondřej Vencálek, Ph.D.

**The year of presentation:** 2018

**Abstract:** This bachelor thesis is dealing with statistical analysis of financial time series, whose values are predicted by statistical models. I will especially use decomposition of time series. The aim of this work is to propose suitable statistical model for concrete financial time series and to compare predicted and real values.

**Key words:** financial time series, statistical model, decomposition, prediction, martingal

**Number of pages:** 66

**Number of appendices:** 0

**Language:** Czech

### Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně pod vedením pana Mgr. Ondřeje Vencálka, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne .....

.....

podpis

## **Poděkování**

Chtěla bych tímto poděkovat Mgr. Ondřeji Vencálkovi, Ph.D. za vedení, cenné rady a připomínky při zpracování mé bakalářské práce.

# Obsah

Úvod	8
<b>1 Časové řady</b>	<b>9</b>
1.1 Druhy časových řad . . . . .	10
1.1.1 Očišťování od kalendářních variací . . . . .	10
1.2 Základní charakteristiky časových řad . . . . .	11
1.2.1 Popisné charakteristiky . . . . .	11
1.2.2 Míry dynamiky . . . . .	12
1.3 Základní přístupy k analýze časových řad . . . . .	12
1.4 Predikce v časových řadách . . . . .	13
<b>2 Finanční časové řady</b>	<b>16</b>
2.1 Míra zisku . . . . .	16
2.2 Lineární nezávislost logaritmů výnosů . . . . .	18
2.3 Leptokurtické rozdělení . . . . .	18
2.4 Nestacionarita . . . . .	18
2.5 Podmíněná heteroskedasticita . . . . .	19
2.6 Volatilita . . . . .	19
<b>3 Klasický (formální) model</b>	<b>21</b>
3.1 Trend v časové řadě . . . . .	22
3.1.1 Lineární trend . . . . .	23
3.1.2 Kvadratický trend . . . . .	24
3.1.3 Exponenciální trend . . . . .	26
3.2 Adaptivní přístupy k modelování časové řady . . . . .	27
3.2.1 Metoda klouzavých průměrů . . . . .	27
3.2.2 Exponenciální vyrovnání . . . . .	28
3.3 Popis sezónní složky . . . . .	31
3.3.1 Periodicita časové řady . . . . .	32
3.4 Analýza náhodné složky . . . . .	33
<b>4 Martingal</b>	<b>35</b>

<b>5</b>	<b>Analýza kurzu CHF/CZK</b>	<b>36</b>
5.1	Základní charakteristika dat . . . . .	37
5.2	Vlastnosti finančních časových řad . . . . .	39
5.2.1	Stacionarita . . . . .	39
5.2.2	Volatilita . . . . .	39
5.2.3	Leptokurtické rozdělení logaritmů výnosů . . . . .	40
5.2.4	Lineární nezávislost logaritmů výnosů . . . . .	42
5.3	Dekompozice . . . . .	42
5.3.1	Analýza trendu . . . . .	42
5.3.2	Analýza periodicity . . . . .	44
5.3.3	Analýza náhodné složky . . . . .	45
5.3.4	Klouzavé průměry . . . . .	46
5.3.5	Exponenciální vyrovnání . . . . .	47
5.4	Predikce . . . . .	49
5.4.1	Model 1 – lineární trend . . . . .	49
5.4.2	Model 2 – exponenciální trend . . . . .	51
5.4.3	Model 3 – exponenciální trend s periodou . . . . .	52
5.4.4	Model 4 – klouzavé průměry . . . . .	53
5.4.5	Model 5 – exponenciální vyrovnání . . . . .	55
5.4.6	Model 6 – martingal . . . . .	57
5.5	Hodnocení predikčních modelů . . . . .	58
5.6	Analýza kurzu CHF/CZK po skončení intervence ČNB . . . . .	60
	<b>Závěr</b>	<b>63</b>
	<b>Literatura</b>	<b>65</b>

# Úvod

Má bakalářská práce Možnosti predikce ve finančních časových řadách se zabývá tématem, které je jednou z nejdůležitějších oblastí v rozvoji statistiky. V ekonomii je modelování časových řad velmi využívanou kvantitativní metodou při analýze dat. Hlavním důvodem analýzy je především konstrukce modelu, která dále umožňuje předpovídat budoucí vývoj systému.

Cílem této bakalářské práce je v první řadě rešerše v oblasti finančních časových řad. Za důležitou část této práce považuji popsání možností predikce ve finančních časových řadách a zaměření se na klasický formální model. Na základě těchto poznatků se pokusím analyzovat finanční časovou řadu, konkrétně kurzu CHF/CZK<sup>1</sup>. Cílem bude tedy analýza chování a vývoje konkrétní finanční časové řady a následná aplikace několika vhodných predikčních modelů na vybraná data. Zhodnocení predikcí je poslením cílem této analýzy.

Práce je rozdělena na část teoretickou a část praktickou. V teoretické části postupně objasním pojmy časová řada, její druhy, možnosti přístupu k její analýze a predikci. Druhá kapitola je zaměřena na typické vlastnosti a předpoklady finančních časových řad. Další kapitola přiblíží jednu z možností přístupu k analýze časových řad a tím je klasický model pomocí dekompozice, kde postupně vysvětlím analýzu všech čtyř složek i následnou konstrukci predikčního modelu. Ve čtvrté kapitole se zmíním o možnosti predikce pomocí martingalu.

V praktické části jsem zvolila analýzu konkrétního měnového kurzu. Data pro tuto práci jsem získala z internetové stránky České národní banky a následně zpracovala pomocí MS Office Excel.

---

<sup>1</sup>CHF - švýcarský frank, CZK - česká koruna



# Kapitola 1

## Časové řady

Tato kapitola vychází zejména z publikací [1], [3], [5], [10], pokud nebude uvedeno jinak.

Možností, jak definovat časovou řadu, je mnoho. Obecně lze říct, že jde o posloupnost pozorování, která jsou uspořádaná v čase. V této práci budeme časovou řadu chápat jako statistickou časovou řadu, jejíž chování je zatíženo nejistotou, na rozdíl od deterministické časové řady, jejíž chování se dá striktně popsat matematickým vzorcem a tudíž přesně zkonstruovat její předpověď.

Časová řada je speciálním případem náhodného, neboli stochastického procesu. Je proto důležité tyto pojmy správně pochopit, což nám umožní následující definice.

### Definice 1.1.

Nechť  $(\Omega, \mathcal{A}, P)$  je pravděpodobnostní prostor, nechť  $T \subset \mathbb{R}$ . Rodina reálných náhodných veličin  $\{y_t, t \in T\}$  definovaných na  $(\Omega, \mathcal{A}, P)$  se nazývá *náhodný proces*.

V případě, že  $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$  nebo v případě  $T = \mathbb{N}_0 = \{0, 1, \dots\}$ , mluvíme o náhodné posloupnosti s diskrétním časem nebo o tzv. *časové řadě*. Pokud  $T \in (a; b)$ , kde  $-\infty < a < b < \infty$ , říkáme, že  $\{y_t, t \in T\}$  je proces se spojitým časem. [4, strana 7]

## 1.1. Druhy časových řad

Je důležité rozlišovat různé druhy časových řad, abychom s nimi mohli dále pracovat nebo je upravovat.

Jako první si uvedeme řadu *stochastickou* a *deterministickou*. U řady deterministické lze s poměrně vysokou přesností predikovat budoucnost, jelikož neobsahují prvek náhody. Naopak řady stochastické, které jsou pro ekonomické problémy typické, jsou náhodou ovlivňovány a jejich predikce je tudíž méně přesná a více nejednoznačná.

Z hlediska okamžiku pozorování existují řady *ekvidistantní* a *neekvidistantní*. Ekvidistantní časové řady mají stejně dlouhé intervaly mezi okamžiky pozorování, naopak neekvidistantní nemají.

Dále existují řady *intervalové* a *okamžikové*, kde v okamžikových časových řadách se hodnoty váží k jednotlivým časovým okamžikům, zatímco intervalové časové řady závisí na délce časového intervalu.

Také je užitečné odlišovat časové řady *dlouhodobé* a *krátkodobé*. Dlouhodobé časové řady mají hodnoty sledované v ročních či delších časových usecích, krátkodobé v intervalu kratším než jeden rok. Speciálním případem krátkodobých časových řad jsou tzv. *vysokofrekvenční časové řady*. Jejich frekvence sledování je velmi častá, může být týdenní, denní či několikrát za den. Do této skupiny patří i finanční časové řady.

### 1.1.1. Očišťování od kalendářních variací

V případě, že máme neekvidistantní časovou řadu z důvodu kalendářních variací, např. různá délka kalendářních měsíců, je vhodné tuto časovou řadu „očistit“ od těchto nežádoucích vlivů, abychom zajistili srovnatelnost.

$$y_t^* = y_t \frac{\bar{k}_t}{k_t},$$

kde  $y_t$  je hodnota očištěvaného ukazatele v příslušném období ( $t = 1, 2, \dots, n$ ).  $k_t$  je počet kalendářních dní v daném měsíci,  $\bar{k}_t$  je počet dnů, tzv. průměrného měsíce.

Kromě problémů týkajících se kalendářních variací existuje řada dalších problémů, se kterými se můžeme v analýze časových řad setkat. Jedná se tedy převážně o problémy s volbou časových bodů pozorování, problémy s nesrovnalostí jednotlivých měření nebo se zvolením vhodné délky časových řad.

## 1.2. Základní charakteristiky časových řad

Základní charakteristiky časových řad nám slouží k hrubé představě o jejím charakteru a chování.

### 1.2.1. Popisné charakteristiky

*Prostý chronologický průměr* se použije v případě, že je délka mezi jednotlivými okamžiky konstantní. Vyjadřuje průměrnou úroveň okamžikového ukazatele.

$$\bar{y}_t = \frac{\left(\frac{y_1+y_2}{2}\right) + \left(\frac{y_2+y_3}{2}\right) + \dots + \left(\frac{y_{n-1}+y_n}{2}\right)}{n-1}.$$

V druhém případě předpokládáme, že délka mezi jednotlivými okamžiky stejná nebude, a proto je nutné jednotlivé dílčí intervaly vážit délkami příslušných intervalů.

*Vážený chronologický průměr* má podobu

$$\bar{y}_t = \frac{\left(\frac{y_1+y_2}{2}d_1\right) + \left(\frac{y_2+y_3}{2}d_2\right) + \dots + \left(\frac{y_{n-1}+y_n}{2}\right)d_{n-1}}{d_1 + d_2 + \dots + d_{n-1}},$$

kde  $y_t$  jsou jednotlivé hodnoty ukazatele a  $d_t$  označuje délky jednotlivých intervalů.

## 1.2.2. Míry dynamiky

### *Absolutní přírůstky*

První diference, které vyjadřují absolutní změny hodnot ukazatelů v čase  $t$  oproti času  $t - 1$ .

$$\Delta Y_t = Y_t - Y_{t-1}, t = 2, \dots, n$$

### *Průměrné přírůstky*

Též je lze nazvat aritmetickým průměrem absolutních přírůstků, který se počítá pro celou časovou řadu. Vyjadřují, o kolik se průměrně změní hodnota ukazatele v jednotlivých okamžicích.

$$\bar{\Delta Y} = \frac{1}{n-1} \sum_{t=2}^n \Delta Y_t = \frac{Y_n - Y_1}{n-1}, t = 2, \dots, n$$

### *Koeficient růstu*

Koeficient růstu nebo též tempo růstu (koeficient růstu vyjádřený v procentech), udává, kolikrát klesla (či vzrostla) hodnota ukazatele v čase  $t$  od předcházejícího období, tj. času  $t - 1$ .

$$k_t = \frac{Y_t}{Y_{t-1}}, t = 2, \dots, n$$

Pokud je koeficient růstu větší než 1, znamená to růst. V opačném případě pokles.

## 1.3. Základní přístupy k analýze časových řad

Existuje mnoho metod pro analýzu časových řad jako například *klasický (formální) model*, který vychází z dekompozice časové řady na čtyři složky, tj. systematické složky (trendovou, sezónní, cyklickou) a náhodnou složku, přičemž pracuje především se složkami systematickými.

Další variantou je *Boxova - Jenkinsonova metodologie*, která při své konstrukci modelu klade důraz na náhodnou složku.

Časovou řadu můžeme také analyzovat pomocí *lineárně dynamických modelů*, které vychází z předpokladu, že časové řady jsou vysvětlovány pomocí hodnot

dalších časových řad, které vysvětlovanou časovou řadu ovlivňují.

Posledním přístupem, který si zde uvedeme, je *spektrální analýza* časových řad, která považuje časovou řadu za směs sinusových a kosinusových křivek s různými amplitudami a frekvencemi.

V této práci se zaměříme zejména na klasický (formální) model.

## 1.4. Predikce v časových řadách

Predikce v časových řadách je jedním z nejdůležitějších úkolů analýzy. Správné porozumění chování časových řad nám může dát kvalitní předpověď budoucnosti, na jejímž základě se můžeme rozhodovat více kvalifikovaněji. Předpovědi mohou být dvojího typu – *bodové* a *intervalové*.

Bodová předpověď nám udává číslo, které vyjadřuje odhad budoucí hodnoty analyzované časové řady. Je dobré si uvědomit, že předpověď je vždy zatížena nějakou chybou. Intervalovou předpověď můžeme brát jako analogii intervalu spolehlivosti ze základů statistiky.

Metody pro vytváření předpovědí rozlišíme na dvě kategorie. *Kvalitativní předpovědi* jsou založeny na názoru experta, který se snaží co nejlépe odhadnout budoucí chování řady. Je zřejmé, že tato metoda je dosti subjektivní. Do této kategorie se řadí i tzv. *Delfi metoda*, která je založena na dotazování určité skupiny odborníků, přičemž každý vyjádří svůj odborný názor. *Kvantitativní předpovědi* jsou založeny na statistických postupech, které jsou daleko objektivnější. Při této metodě však musíme předpokládat, že se charakter časové řady nebude v budoucnu měnit. V této práci se zaměříme na kvantitativní metody předpovědí.

Výběr příslušné předpovědní techniky záleží na mnoha aspektech. Určitou roli hraje především požadovaná forma a přesnost předpovědí, horizont předpovědi, charakter zkoumaných dat či jejich dostupnost.

Odhady náhodné složky, neboli *reziduum*  $\epsilon_t$  se odhadne následujícím vztahem

$$\hat{\epsilon}_t = y_t - \hat{y}_t,$$

přířemž  $y_t$  jsou reálné hodnoty a  $\hat{y}_t$  jsou hodnoty našeho odhadnutého modelu. Hlavním důvodem chybovosti je zejména reziduální, tj. náhodná složka, která představuje náhodnou fluktuaci v datech. K posouzení kvality modelů nám slouží míry kvality.

Poznámka. Pokud chceme ohodnotit kvalitu predikcí, stejným způsobem jako reziduum vypočteme tzv. *predikční chyby*.

Nejčastějším hodnocením kvality je součet čtvercových chyb SSE (Sum of Squared Errors), který vyjadřuje odchylky skutečných hodnot a modelu odpovídajících hodnot. Čím menší tato hodnota bude, tím přesnější je model.

$$SSE = Se = \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \sum_{t=1}^n \epsilon_t^2.$$

SSE s počtem pozorování samozřejmě roste, po jejím zprůměrování však získáme tzv. střední čtvercovou chybu MSE (Mean square error).

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

Koeficient determinace nám udává, jakou část variability v pozorováních jsme vysvětlili modelem.

$$R^2 = 1 - \frac{Se}{St}, \quad R^2 \in [0, 1],$$

kde  $Se$  je součet čtvercových chyb a  $St$  lze chápat jako celkovou variabilitu v datech, lze ji vyjádřit jako  $\sum_{t=1}^n (y_t - \bar{y})^2$ . Pochopitelně chceme, aby  $R^2$  bylo co největší.

Poznámka. Při posuzování míry vhodnosti modelu není vhodné toužit pouze po vysokém koeficientu determinace, či nízkému reziduálnímu součtu čtverců. Je zřejmé, že bychom vždy volili polynom vyššího stupně, který by dokonale proložil historické hodnoty, ale mohl by selhat při predikci. Cílem totiž není dokonalé proložení historických hodnot, ale zjednodušení a vystihnutí základního trendu a případně dalších systematických složek. Proto se snažíme najít model, který bude jednoduchý, ale věrný. Při rozhodování o stupni polynomu je vhodné

použít testy hypotéz. Statistický test hypotéz lze možno provést například pomocí kritického oboru, pomocí intervalu spolehlivosti či p-hodnoty.

# Kapitola 2

## Finanční časové řady

Finanční trh je místo, na němž se setkávají nabídky peněžních prostředků v různých měnách, dlouhodobých cenných papírů a akcií, s poptávkami po nich. Finanční časové řady podávají kvantitativní informace o finančním trhu. Ceny výše uvedených forem kapitálu jsou sledovány v určitém období s určitou frekvencí a tvoří časovou řadu. Finanční časové řady jsou speciálním případem ekonomických řad. Jsou typické především vysokou frekvencí pozorování, proto se jim říká *vysokofrekvenční*. Finanční data mají svůj specifický charakter. Obvykle v nich bývá obtížné najít systematické složky z důvodu výrazného zastoupení šumové složky, nemívají normální rozdělení a v jejich tvaru se projevuje chování a způsob fungování finančních trhů.

V této kapitole bych se ráda věnovala jednotlivým vlastnostem jako je například nestacionarita, shlukování volatility, podmíněná heteroskedasticita a v neposlední řadě i leptokurtické rozdělení.

Celá kapitola vychází převážně z publikací [2], [4], [5], [8], [9], pokud nebude uvedeno jinak.

### 2.1. Míra zisku

Cenové změny, které sledujeme, mohou být definovány různě. Uvedeme si nejčastější, a tou je relativní cenová změna nebo také míra zisku. Míra zisku nám



popisuje rozdíl mezi cenou v čase  $t$ , tj. na konci uvažovaného období a v čase  $t - 1$ , tj. na počátku uvažovaného období. Tento rozdíl je relativně vztažen k ceně na počátku uvažovaného období.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Cena aktiva nemůže být záporné číslo. To znamená, že minimální dosažitelný relativní přírůstek ceny, také se mu říká minimální dosažitelný jednoduchý výnos aktiva, nabývá hodnoty  $-1$ , jelikož platí vztah :

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \geq -1.$$

Při tomto vyjádření máme problém, kdy cena aktiva  $P_t$  může nabývat jakéhokoli reálného čísla, tudíž nemá horní ani spodní mez. Tento problém můžeme vyřešit následující úvahou. Jestliže jednoduché výnosy definujeme jako

$$R_t + 1 = \frac{P_t}{P_{t-1}},$$

měly by mít rozdělení nezáporné náhodné veličiny. V tom případě lze hodnoty těchto časových řad generovat logaritmicke-normálním rozdělením. Logaritmus náhodné veličiny s logaritmicke-normálním rozdělením však má rozdělení normální.

$$r_t = \ln(R_t + 1) = \ln P_t - \ln P_{t-1}.$$

Tuto logaritmicke míru zisku chápeme jako absolutní změnu logaritmickeých cen z času  $t - 1$  do času  $t$ . V případě malé výnosnosti ( $R_t \rightarrow 0$ ), lze pomocí Taylorova rozvoje <sup>1</sup> aproximovat

$$r_t = \ln(R_t + 1) \sim R_t,$$

a považvat obě míry zisku za ekvivalentní.

---

<sup>1</sup>viz [6]

## 2.2. Lineární nezávislost logaritmů výnosů

Při analýze finančních časových řad se předpokládá, že logaritmy výnosů jsou nekorelované a nezávislé<sup>2</sup>. Nekorelovanost v mírách zisku může být splněna, neznamená to však jejich nezávislost obecně.

## 2.3. Leptokurtické rozdělení

Finanční časové řady obvykle vychází z předpokladu, že logaritmy výnosů mají normální rozdělení se střední hodnotou  $\mu$  a konstantním rozptylem  $\sigma_r^2$ , tj.  $r_t \sim N(\mu, \sigma_r^2)$ . Normální rozdělení je symetrické, což znamená, že koeficient šikmosti  $SK_r$ , který udává, zda jsou hodnoty kolem zvoleného středu rozloženy souměrně, nabývá hodnoty nula. Naopak koeficient špičatosti  $K_r$ , který měří stupeň koncentrace hodnot kolem středu, je roven číslu 3. Tyto charakteristiky jsou definovány vztahy

$$SK_r = E \left[ \frac{(r_t - \mu)^3}{\sigma_r^3} \right] = 0, \quad K_r = E \left[ \frac{(r_t - \mu)^4}{\sigma_r^4} \right] = 3.$$

Pro finanční data je však typické tzv. *Leptokurtické rozdělení*, které je velmi podobné rozdělení normálnímu. Finanční data se ale oproti normálnímu rozdělení vyznačují „těžšími konci“ (fat tails) a zároveň „tenkým pasem“ (thin waist). To znamená, že náhodná veličina  $r_t$  má větší četnost jak pozorování velmi vzdálených od střední hodnoty, tak větší četností hodnot blízkých střední hodnotě. Koeficient špičatosti je v případě Leptokurtického rozdělení větší než tři.

## 2.4. Nestacionarita

Nestacionarita finančních časových řad je jednou z dalších typických vlastností. Jak již plyne z názvu, jde o časové řady, které nesplňují podmínky stacionarity, tj. mění v čase svou kovarianční strukturu.

---

<sup>2</sup>V případě normálního rozdělení tyto dvě vlastnosti splývají.

## 2.5. Podmíněná heteroskedasticita

Společnou vlastností finančních časových řad je také proměnlivá variabilita v čase, která se může měnit někdy ve velmi krátkých úsecích, jindy ve shlučích.

K detekci heteroskedasticity obvykle stačí pouhé vykreslení reziduí do grafu v závislosti na proměnné.

Pro časové řady měr zisku bývají nejvhodnější modely, které se snaží dynamicky vysvětlit rozptyl, který je podmíněný informací o předchozím chování řady. Tomuto rozptylu říkáme *podmíněný rozptyl* a z toho plyne i název podmíněná heteroskedasticita.

## 2.6. Volatilita

Volatilitu můžeme chápat jako náhodné změny rozptylu v čase. Ve finančních datech je především spojována s nejistotou na trhu. Její vlastností je stacionarita, což znamená, že kolísá kolem svého dlouhodobého průměru a tudíž nediverguje k nekonečnu. Typickou vlastností finančních časových řad bývá shlukování volatility, kdy se střídají období s nízkou volatilitou s obdobími s vysokou volatilitou. Předpokládáme tedy, že se určitá míra volatility pro učitě období drží stejné úrovně. Znamená to tedy, že míra volatility v čase  $t$  závisí na volatilitě v čase  $(t - 1)$ .

S volatilitou obecně souvisí pojem *pákový efekt*. Jedná se o efekt, který způsobuje rapidnější zvětšení volatility po cenovém poklesu nežli po cenovém nárůstu. Tento efekt se však nedá popsat lineárními modely.

Nejjednodušším přístupem je výpočet směrodatné odchylky, která variabilitu měří. Směrodatná odchylka, neboli historická volatilita, je definována vztahem

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}},$$

kde  $\bar{x}$  značí aritmetický průměr. Tento model se považuje za nejstarší přístup k modelování volatility, proto nese název historická volatilita. V dnešní době se používá především ke stanovení tzv. *benchmarks*, neboli srovnávacích hodnot,

k porovnání efektivnosti komplexnějších modelů volatility.

*EWMA-model volatility* je dalším možným přístupem k volatilitě. Dá se interpretovat jako jednoduché exponenciální vyrovnání pro volatilitu, které si vysvětlíme ve třetí kapitole.

K modelování volatility lze ovšem použít i řadu dalších přístupů, jako například autoregresní modely, ARCH modely, GARCH modely či například pomocí implikované volatility. Jejich cíl je vždy stejný a to, co nejlépe analyzovat vývoj volatility v minulosti, z důvodu přesnějšího určení nejistoty v predikcích budoucích hodnot.

Proměnlivá volatilita hodnot výnosů může souviset s úrovní a silou autokorelace. Její přítomnost můžeme testovat například Durbinovým-Watsonovým testem autokorelovanosti.

# Kapitola 3

## Klasický (formální) model

Při zpracování této kapitoly jsem čerpala z publikací [1], [3], [5], [12].

Nejjednodušším a zároveň nejužívanějším konceptem modelování časových řad je model jednorozměrný :

$$y_t = f(t, \epsilon_t), \quad t = 1, 2, \dots, n,$$

kde  $y_t$  je modelová hodnota ukazatele v čase  $t$ ,  $f$  je nějaká známá elementární funkce a  $\epsilon_t$  nazýváme náhodným kolísáním nebo též náhodnou odchylkou. Je dobré zdůraznit, že předpokládáme nekorelovanost náhodných odchylek v různých časech.

Můžeme se setkat i s modely vícerozměrnými, které vycházejí z předpokladu, že vývoj analyzovaného ukazatele není ovlivňován pouze časem, ale i řadou dalších ukazatelů. My se v této práci zaměříme na modely jednorozměrné.

Předpokladem pro klasický (formální) model časových řad, dále jen klasický model, je fakt, že časová řada závisí pouze na čase. Jak jsem již zmínila v úvodu, vychází z rozkladu řady na čtyři složky, pomocí kterých se snaží identifikovat chování jednotlivých složek zvlášť, nikoli celé časové řady najednou.<sup>1</sup> Rozdělení složek je následující:

- Systematické složky
  - **Trend**  $T_t$  - základní tendence vývoje,

---

<sup>1</sup>Pozn. Časová řada nemusí obsahovat všechny složky.

- **Sezónnost**  $S_t$  - pravidelné, tj. periodické odchylky od trendu s periodou menší než 1 rok,
- **Cyklus**  $C_t$  - dlouhodobé kolísání okolo trendu s periodou větší než jeden rok,
- **Náhodná složka**  $\epsilon_t$  - odchylky nevysvětlené pomocí předchozích tří složek.

Základem pro tento přístup je modelování nenáhodné složky. Dekompozice může mít dvojí tvar:

*Aditivní*, při kterém mají všechny složky stejnou jednotku

$$y_t = T_t + C_t + S_t + \epsilon_t = Y_t + \epsilon_t,$$

kde  $Y_t$  vyjadřuje modelovou složku.

*Multiplikativní*, ve které má pouze trendová složka stejnou jednotku jako  $y_t$ , ostatní složky jsou bezrozměrné:

$$y_t = T_t \cdot C_t \cdot S_t \cdot \epsilon_t.$$

### 3.1. Trend v časové řadě

Trend, neboli základní tendence vývoje časové řady, se dá popsat jednoduchým způsobem a tím je vyrovnaní hodnot časové řady matematickou funkcí. Touto eliminací trendové složky od složek ostatních získáme úplnou informaci o hlavní tendenci vývoje v čase a budeme moci předpovídat i její budoucí vývoj. Kromě klasických postupů eliminace jako je proložení hodnot funkcí lineární, kvadratickou, exponenciální aj., lze použít při konstrukci také adaptivní přístupy jako jsou klouzavé průměry či exponenciální vyrovnaní.

V této práci si z klasických postupů přiblížíme převážně trend lineární, kvadratický a exponenciální, a to z důvodu užití v praktické části této bakalářské práce.

Při analýze trendu časové řady obvykle předpokládáme tvar

$$y_t = T_t + \epsilon_t.$$

Z rozmanité nabídky trendových funkcí vybíráme v první řadě subjektivně na základě grafického znázornění dat nebo z předpokládaného chování dat, například z ekonomického hlediska.

Zaměříme se na metody odhadu parametrů trendových funkcí, které jsou lineární v parametrech. Můžeme k nim tudíž použít nejznámější metodu a tou je metoda nejmenších čtverců viz [5, str.34].

### Předpoklady lineárního regresního modelu

- $E(\epsilon_t) = 0, \quad \forall t,$
- $var(\epsilon_t) = \sigma^2, \quad \forall t,$
- $cor(\epsilon_t, \epsilon_{t^*}) = 0, \quad \forall t \neq t^*,$
- $\epsilon_t \sim N(0, \sigma^2), \quad \forall t$
- regresní parametry  $\beta$  mohou nabývat libovolných hodnot,
- regresní model je lineární v parametrech.

#### 3.1.1. Lineární trend

V případě lineárního trendu, předpokládáme, že hodnoty lze proložit přímkou, tj.

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

Parametr  $\beta_0$  vyjadřuje střední hodnotu sledované veličiny v čase  $t = 0$  a parametr  $\beta_1$  můžeme interpretovat jako průměrný přírůstek za jednu časovou jednotku.

Jak již bylo uvedeno v úvodu této kapitoly, parametry můžeme odhadnout metodou nejmenších čtverců, kdy minimalizací následujícího výrazu dokážeme vyjádřit odhady.

$$\min \sum_{t=1}^n (y_t - (\beta_0 + \beta_1 t))^2$$

Dostaneme soustavu rovnic

$$\sum_{t=1}^n y_t = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{t=1}^n t$$

$$\sum_{t=1}^n ty_t = \hat{\beta}_0 \sum_{t=1}^n t + \hat{\beta}_1 \sum_{t=1}^n t^2$$

Přičemž řešením této soustavy docházíme k odhadům parametrů  $\beta_0$  a  $\beta_1$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{t}, \quad \hat{\beta}_1 = \frac{\sum_{t=1}^n ty_t - \bar{t} \sum_{t=1}^n y_t}{\sum_{t=1}^n t^2 - n\bar{t}^2},$$

kde

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

Chceme-li provést bodovou predikci očekávaných budoucích hodnot  $y_T$ , vypočítáme ji pouhým dosazením času  $T$ , na který chceme predikci vypočítat, do trendové přímky, kterou jsme si právě odhadli

$$\hat{y}_T = \hat{\beta}_0 + \hat{\beta}_1 T.$$

Intervalový predikční odhad vypočítáme vztahem

$$(\hat{\beta}_0 + \hat{\beta}_1 T) \pm t_{(n-2), (1-\frac{\alpha}{2})} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(T - \bar{t})^2}{\sum_{t=1}^n t - n\bar{t}^2}},$$

kde  $t_{(n-2), (1-\frac{\alpha}{2})}$  je  $(1 - \frac{\alpha}{2}) \cdot 100$  procentní kvantil  $t$ -rozdělení o  $(n - 2)$  stupních volnosti a  $s$  vypočteme vztahem

$$s = \sqrt{\frac{1}{n-2} \sum_{t=1}^n (y_t - \hat{y}_t)^2}.$$

### 3.1.2. Kvadratický trend

Kvadratický trend má tvar

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2,$$



odtud opět pomocí metody nejmenších čtverců minimalizujeme výraz

$$\min \sum_{t=1}^n [y_t - (\beta_0 + \beta_1 t + \beta_2 t^2)]^2,$$

dostaneme soustavu rovnic

$$\begin{aligned} \hat{\beta}_0 n + \hat{\beta}_1 \sum_{t=1}^n t + \hat{\beta}_2 \sum_{t=1}^n t^2 &= \sum_{t=1}^n y_t, \\ \hat{\beta}_0 \sum_{t=1}^n t + \hat{\beta}_1 \sum_{t=1}^n t^2 + \hat{\beta}_2 \sum_{t=1}^n t^3 &= \sum_{t=1}^n t y_t, \\ \hat{\beta}_0 \sum_{t=1}^n t^2 + \hat{\beta}_1 \sum_{t=1}^n t^3 + \hat{\beta}_2 \sum_{t=1}^n t^4 &= \sum_{t=1}^n t^2 y_t. \end{aligned}$$

Řešením této soustavy dostaneme odhad parametrů  $\beta_0$ ,  $\beta_1$  a  $\beta_2$ .

Předpověď budoucí hodnoty  $y_T$  má potom tvar

$$\hat{y}_T = \hat{\beta}_0 + \hat{\beta}_1 T + \hat{\beta}_2 T^2.$$

Intervalovou předpověď lze vyjádřit vztahem

$$\hat{y}_T \pm t_{(n-3), (1-\frac{\alpha}{2})} \cdot s \cdot \sqrt{1 + \begin{pmatrix} 1 & T & T^2 \end{pmatrix} (X'X)^{-1} \begin{pmatrix} 1 \\ T \\ T^2 \end{pmatrix}},$$

kde opět  $t_{(n-3), (1-\frac{\alpha}{2})}$  je  $(1 - \frac{\alpha}{2}) \cdot 100$  procentní kvantil t-rozdělení o  $(n-3)$  stupních volnosti a dále platí

$$s = \sqrt{\frac{1}{n-3} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \dots & \dots & \dots \\ 1 & n & n^2 \end{pmatrix}$$

Poznámka.  $X'$  značí transpozici matice  $X$ .

### 3.1.3. Exponenciální trend

Exponenciální trend má obecně tvar

$$T_t = \alpha\beta^t.$$

$\alpha$  zde vyjadřuje hodnotu trendu v čase 0 a  $\beta$  můžeme interpretovat jako tempo růstu, které je konstantní v čase. Jestliže je  $\alpha > 0$ , potom v případě  $\beta > 1$  dochází k růstu a v případě  $0 < \beta < 1$  nastává pokles.

Jelikož exponenciální funkce v obecném tvaru není lineární vzhledem k parametrům, musíme v prvním případě provést lineární transformaci původní trendové exponenciály pomocí logaritmu, následně můžeme použít metodu nejmenších čtverců k odhadu parametrů  $\hat{\alpha}$  a  $\hat{\beta}$ .

$$\ln(T_t) = \ln(\alpha) + t \cdot \ln(\beta).$$

$$\hat{\alpha} = e^{\ln\hat{\alpha}}$$

$$\hat{\beta} = e^{\ln\hat{\beta}}$$

$$\hat{y}_t = e^{\ln\hat{T}_t}$$

Predikci budoucích hodnot potom lze vypočítat vztahem

$$\hat{y}_T = \hat{\alpha}\hat{\beta}^T$$

K odhadu parametrů lze využít i jiné metody. Například metodu vážených nejmenších čtverců po logaritmické transformaci nebo metodu vybraných bodů, která je vhodná především k hrubému odhadu, jelikož není příliš přesná.

## 3.2. Adaptivní přístupy k modelování časové řady

Všechny trendy, které jsme si doposud přiblížili, předpokládaly, že se v průběhu nemění jejich charakter, a proto se jim také někdy říká modely s neměnnými parametry.

Adaptivní přístupy k trendové složce mají tu schopnost, že umějí pracovat s trendy, které v čase mění svůj charakter. Dá se tedy říct, že nepředpokládají stabilitu modelu, tudíž se nedají popsat matematickou křivkou s neměnnými parametry. V obecném případě nemusí být splněna ani spojitost trendové funkce.

Použijeme-li tedy křivku s měnivými parametry, časovou řadu můžeme vyrovnat v kratších úsecích, kdy bude mít křivka v každém úseku jiné parametry. Obecně mluvíme o lokálním vyrovnání trendu neboli koncepci postupného trendu. Praktickou výhodou těchto metod je konstrukce přesnějších předpovědí, jelikož reagují na časové změny charakteru dané řady.

### 3.2.1. Metoda klouzavých průměrů

Jako *metodu klouzavých průměrů* označujeme lineární kombinaci členů řady původní. Podstata tohoto modelu je, že posloupnost původních empirických pozorování nahradíme řadou průměrů, které jsou z nich vypočítané. Název této metody vychází z postupného výpočtu průměrů, vždy o jedno pozorování dopředu - tj. „kloužeme“. Důležité je stanovení si počtu pozorování, tzv. délky okna, z nichž jsou jednotlivé průměry počítány. Obecně platí, že čím větší je délka daného okna, tím více je řada vyhlazená. Při volení délky rozhodujeme především subjektivně, podle charakteru dat, důraz je kladen především na sezónní a periodické fluktuace, které se snažíme vyhladit.

Můžeme se setkat s různými typy klouzavých průměrů. Existují *centrované* nebo naopak *necentrované* klouzavé průměry, které se dále v obou případech dělí na *prosté* a *vážené*. Při necentrovaných klouzavých průměrech se bavíme o délce okna lichého počtu ( $2m + 1$ ), což znamená, že střed okna odpovídá jednomu z uvažovaných pozorování. Centrované klouzavé průměry jsou o malinko složitější, a to z důvodu centrování. Délka okna je sudá ( $2m$ ), a tudíž střed okna neodpovídá

žádnému pozorování. Tento typ je vhodný v případě, že má perioda časové řady sudý počet časových jednotek. Co se týče prostých klouzavých průměrů, data prokládáme lineárním trendem a při vážených klouzavých průměrech polynomem vyššího řádu.

Je dobré si uvědomit, že použitím této metody zůstane prvních a posledních  $m$  hodnot nevyrovnaných. Tento problém se dá řešit například proložením koncového okna vhodným polynomem, pomocí kterého lze následně predikovat budoucí hodnoty dané řady.

Alternativou klouzavých průměrů jsou klouzavé mediány, které dokáží z řady vyloučit odlehlá pozorování. Postup je analogický.

### 3.2.2. Exponenciální vyrovnání

Hned na úvod je nutné zdůraznit, že exponenciální vyrovnání nemá žádnou spojitost s proložením časové řady exponenciálou. Jedná se však o jeden z nejčastějších způsobů vyhlazování a předpovídání časové řady. Jeho výhodou oproti metodě klouzavých průměrů je skutečnost, že nevolíme délku vyhlazovacího okna, protože výpočet vyrovnaných hodnot je založen na váženém průměru všech hodnot předchozích pozorování. Vyrovnaná řada  $\hat{y}_t$  pak minimalizuje výraz

$$\min \sum_{j=0}^n (y_{t-j} - \hat{y}_{t-j})^2 \cdot \alpha^j,$$

kde  $\alpha \in (0, 1)$ , a jedná se o vyrovnávací konstantu. Z tohoto výrazu také vyplývá, že váhy exponenciálně klesají, což znamená, že vzdáleným pozorováním přiřazujeme čím dál tím nižší váhy.

Za nejužívanější se považuje *jednoduché exponenciální vyrovnání*, *dvojitě exponenciální vyrovnání* nebo jeho zobecnění v podobě *Holtovy metody*, kterou se v této práci nebudeme zabývat.

#### Jednoduché exponenciální vyrovnání

Základním předpokladem jednoduchého exponenciálního vyrovnání je přibližně konstantní trend časové řady. Hodnota trendu se tudíž rovná hodnotě parame-

tru  $\beta_0$ .

$$T_t = \beta_0$$

Označíme-li jako  $\hat{\beta}_0(t)$  odhad parametru  $\beta_0$  provedený v čase  $t$ . Odhad bude představovat jak predikci budoucích hodnot, tak vyrovnanou hodnotu  $\hat{y}_t$ . Získáme jej minimalizací výrazu :

$$\min \sum_{j=0}^{\infty} (y_{t-j} - \beta_0)^2 \alpha^j.$$

Tento výraz zderivujeme podle  $\beta_0$  a následně derivaci položíme rovnu nule. Při snadné úpravě dostaneme rekurentní vyjádření odhadu parametru  $\beta_0$  a zároveň vyrovnanou hodnotu  $\hat{y}_t$ .

$$\hat{y}_t = (1 - \alpha)y_t + \alpha\hat{y}_{t-1}.$$

Aktuální odhad trendu je tudíž váženým průměrem předešlého odhadu trendu a aktuálního pozorování, přičemž vyrovnávací konstanta  $\alpha$  určuje rozdělení vah v tomto průměru. Čím menší  $\alpha$  zvolíme, tím rychleji metoda reaguje na změny v charakteru dat a klade tak důraz na aktuální pozorování. V opačném případě se naopak zesílí vyrovnávací schopnost metody. Při volbě  $\alpha$  se můžeme řídit několika doporučeními, např.  $\alpha \in (0, 7; 1)$ .

Pro predikci pak používáme následující vztah

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

kde  $\tau > 0$  je libovolné, přičemž  $\hat{y}_{t+\tau}(t)$  vyjadřuje predikci hodnoty  $y_{t+\tau}$  zkonstruovanou v čase  $t$ .

Poznámka. Při určení počáteční hodnoty  $\hat{y}_0$  se doporučuje vypočítat průměr z prvních  $n = 6$  hodnot pozorování nebo z první poloviny dat, tj.

$$\hat{y}_0 = \frac{1}{n} \sum_{t=1}^n y_t, \quad n = 6.$$

## Dvojitě exponenciální vyrovnání

Dvojitě exponenciální vyrovnání lze použít v případě, že trend analyzovaných dat můžeme v krátkých usecích považovat za lineární, tj.  $T_t = \beta_0 + \beta_1 t$ .

Odhady parametrů  $\beta_0$  a  $\beta_1$  v čase  $t$ , které označujeme  $\hat{\beta}_0$  a  $\hat{\beta}_1$ , lze získat minimalizací výrazu

$$\min \sum_{j=0}^{\infty} [y_{t-j} - (\beta_0 + \beta_1(-j))]^2 \alpha^j,$$

kde  $t$  vyjadřuje časový počátek soustavy souřadnic. Tento výraz zderivujeme podle obou parametrů, derivace položíme rovny nule a dostaneme soustavu rovnic, kterou po vyřešení upravíme na co nejjednodušší tvar, a zavedeme následující veličiny.

*Jednoduchá vyrovnávací statistika*  $S_t$  s předpisem

$$S_t = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j y_{t-1} = (1 - \alpha)y_t + \alpha S_{t-1},$$

kde  $S_t$  odpovídá hodnotě jednoduché vyrovnávací statistiky.

*Dvojitá vyrovnávací statistika*  $S_t^{[2]}$  je definovaná jako

$$S_t^{[2]} = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j S_{t-j} = (1 - \alpha)S_t + \alpha S_{t-1}^{[2]}.$$

Při drobných úpravách dostaneme hledané odhady parametrů

$$\hat{\beta}_0(t) = 2S_t - S_t^{[2]}, \quad \hat{\beta}_1(t) = \frac{1 - \alpha}{\alpha}(S_t - S_t^{[2]}).$$

Predikci v čase  $t$  pro čas  $t + \tau$ , kde  $\tau > 0$  můžeme vyjádřit vztahem

$$\hat{y}_{t+\tau}(t) = \hat{\beta}_0(t) + \hat{\beta}_1(t)\tau.$$

### 3.3. Popis sezónní složky

V této části práce se zaměříme na sezónnost a možnosti, jak ji analyzovat, abychom měli větší znalost o chování určitého ekonomického jevu. Tato znalost nám pomůže především k přesnější konstrukci predikcí. Jak jsme si již řekli, za sezónnost považujeme periodicky se opakující kolísání hodnot v časové řadě. Je však nutné zdůraznit, že perioda nesmí přesáhnout jeden rok, v tom případě bychom to museli považovat za cyklus. Další zavádějící věcí může být délka periody vzhledem k frekvenci pozorování, která musí být vždy vyšší. Můžeme říct, že sezónní složka tzv. reprezentuje sezónní vlivy, kterými rozumíme soubor přímých a nepřímých příčin, které se během roku opakují.

Pokud budeme předpokládat reálnou existenci sezónních výkyvů, tj. sezónní výkyvy považujeme za statisticky významné, musíme je v prvním případě kvantifikovat. Dalším důležitým cílem je tzv. sezónní očišťování. Jako výstup dostaneme časovou řadu, ve které bude odstraněna nebo alespoň potlačena sezónní složka (obvykle včetně reziduální složky). Toto očištění nám umožní provádět efektivnější a kvalifikovanější studium dlouhodobých tendencí, kterým je časová řada podřízena.

Při metodách klouzavých průměrů jsme se sice snažili potlačit především reziduální složku, dá se však předpokládat, že mohlo dojít k samovolnému částečnému očištění sezónní a cyklické složky. Kromě klasických klouzavých průměrů byly vyvinuty i speciální klouzavé průměry, které jsou upraveny právě tak, aby mohly časovou řadu dokonale, ne jen částečně, sezónně očistit. Jako příklad si můžeme uvést metodu Census X-11 <sup>2</sup>.

Modely sezónnosti mohou mít různé podoby, především z pohledu rozdílného charakteru trendu analyzované časové řady a také vzájemného vztahu trendové a sezónní složky. Ze statistického hlediska může nastat sezónnost dvojího druhu. *Sezónnost konstantní*, kdy efekt sezónny je pro každý rok stejný, tzn. amplituda sezónnosti se nemění v závislosti na trendové složce. Druhým případem je

---

<sup>2</sup>Tato metoda je detailněji popsána např. Shiskin, Young, Musgrave: The X – 11 variant of the Census Method II. seasonal adjustment program.

sezónnost *proporcionální*, kdy je efekt sezónnosti přímo úměrný hodnotě trendu.

### 3.3.1. Periodicita časové řady

Jesliže časová řada obsahuje cyklickou či sezónní složku, můžeme hovořit o *periodické složce*  $P_t$ . Tu modelujeme jako součet sinusoid a kosinusoid o různých frekvencích a amplitudách. Tento model nám umožňuje provést explicitní popis periodicity časové řady. Hovoříme zde o tzv. *spektrální (harmonické) analýze*, přičemž jejím základem je tzv. *model skrytých period*

$$y_t = a_0 + \sum_{j=1}^H a_j \sin(\omega_j t) + \sum_{j=1}^H b_j \cos(\omega_j t), \quad t = 1, \dots, n,$$

kde  $\omega_j = \frac{2\pi j}{n}$  vyjadřuje  $j$ -tou frekvenci z intervalu  $(0, \pi)$ ,  $H = \frac{n}{2}$  pro  $n$  sudé a  $H = \frac{n-1}{2}$  pro  $n$  liché a vyjadřuje maximální počet period.  $a_0, a_j, b_j \in \mathbb{R}$  jsou regresní parametry modelu, které odhadneme pomocí metody nejmenších čtverců následovně:

$$a_0 = \frac{1}{n} \sum_{t=1}^n y_t = \bar{y}_n,$$

$$a_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(\omega_j t), \quad j = 1, \dots, H,$$

$$b_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(\omega_j t), \quad j = 1, \dots, H.$$

Poznámka. Předpokládáme časovu řadu v aditivním tvaru s konstantním trendem.

K analýze, které periody jsou důležité, nám slouží *periodogram*. Pro všechny frekvence  $\omega_j$  vypočteme

$$I_j = \frac{1}{2}(\hat{a}_j^2 + \hat{b}_j^2), \quad j = 1, \dots, H.$$

Z periodogramu lze sice zjistit, které periody jsou statisticky důležité a které nikoli, toto vyhledávání vrcholů je však poměrně subjektivní. V praxi se proto



užívá řada testů.

Jedním z testů je Fisherův test periodicity, kde testujeme nulovou hypotézu

$$H_0 = y_t = a_0 + \epsilon_t,$$

proti alternativě, že v modelu skrytých period existuje alespoň jedna významná perioda. Hodnoty  $I_j$  si uspořádáme dle velikosti, tj  $I_{[1]} \geq I_{[2]} \geq \dots \geq I_{[H]}$ . Testová statistika  $W$  potom má tvar

$$W = \frac{\max I_j}{\sum_{j=1}^H I_j}.$$

Kritický obor na hladině významnosti  $\alpha$ , obvykle  $\alpha = 0,05$ , s kritickou hodnotou tohoto testu  $g_\alpha$ , která je tabelovaná, pak srovnáváme s námi vypočtenou statistikou. Pokud platí  $W \geq g_\alpha$ , pak zamítáme nulovou hypotézu a našli jsme frekvenci periody. Test opakujeme pro případné nalezení dalších period.

### 3.4. Analýza náhodné složky

Jak jsme si již dříve řekli, náhodná složka má typické předpoklady, jako je její nulová střední hodnota, konstantní rozptyl, nezávislost a předpoklad stejného normálního rozdělení. K ověření právě těchto předpokladů o náhodné složce nám slouží analýza reziduí. Ta je založena na testu nulové hypotézy  $H_0 : \epsilon_t \sim iid$ , přičemž zkratka *iid* vyplývá z anglického „independent and identically distributed random variables“ a znamená nezávislost a stejné rozdělení  $\epsilon_t$ . Analýza reziduí má řadu testů, které lze použít, avšak my si přiblížíme pouze jeden, a tím bude *test založený na znaménkách diferencí*.

Jak již název říká, tento test je založen na znaménkách prvních diferencí reziduí  $V_t$ . Přesněji řečeno – na počtu kladných rozdílů, tj. bodech růstu, který si označíme jako  $S$ . V případě, že je hodnota dvou po sobě jdoucích reziduí stejná, jedno se vynechá.

Náhodná veličina  $V_t$  je potom definována jako

$$V_t = \begin{cases} 1 & \hat{\epsilon}_t - \hat{\epsilon}_{t-1} > 0, \quad t = 2, \dots, n, \\ 0 & \hat{\epsilon}_t - \hat{\epsilon}_{t-1} < 0, \quad t = 2, \dots, n. \end{cases}$$

Potom

$$S = \sum_{t=2}^n V_t.$$

V praxi se dále používá tato asymptotická verze příslušného testu s kritickým oborem

$$\frac{|S - \frac{n-1}{2}|}{\sqrt{\frac{n+1}{12}}} \geq u_{1-\frac{\alpha}{2}},$$

kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2}) \cdot 100$  procentní kvantil normálního rozdělení  $N(0, 1)$ . Pokud daná nerovnost platí,  $H_0$  zamítáme.

# Kapitola 4

## Martingal

K analýze finančních časových řad neodmyslitelně patří teorie efektivního trhu. Ta se zabývá především vývojem cenných papírů, speciálně zaměřeným na akcie. Teorie předpokládá, že kurzy akcií zahrnují očekávání a informace od všech účastníků trhu, tudíž jsou jejich změny nepredikovatelné. Martingal se dá považovat za jeden z nejstarších modelů „chování“ cen akcií, obecně aktiv. Dle [2] jej můžeme interpretovat následovně.

Předpokládejme, že  $P_t$  představuje cenu aktiva v čase  $t$ , potom předpověď ceny aktiva v čase  $t + 1$  se rovná ceně aktiva v čase  $t$ .

$$\hat{P}_{t+1} = P_t.$$

Model nám tedy říká, že za nejlepší, alespoň podle SSE, předpověď zítřejší ceny se dá považovat cena dnešní. Předpokladem je znalost všech cen aktiva v minulosti.

Martingal lze také vyjádřit vztahem

$$P_t = P_{t-1} + a_t,$$

kde  $a_t$  značí přírůstek martingalu a předpokládá se, že  $\{a_t\}$  je proces bílého šumu<sup>1</sup>, který má kromě předpokladu nekorolovanosti náhodných veličin také předpoklad, že jsou veličiny stejně rozdělené s nulovou střední hodnotou a konstantním rozptylem.

---

<sup>1</sup>Bílým šumem nazýváme posloupnost nekorolovaných náhodných veličin  $\epsilon_t$ , které mají nulové střední hodnoty a konstantní rozptyl.

# Kapitola 5

## Analýza kurzu CHF/CZK

V této kapitole se pokusím analyzovat finanční časovou řadu kurzu měny. Pro tuto analýzu jsem vybrala data denního kurzu CHF/CZK <sup>1</sup> za rok 2017. Toto období jsem zvolila z důvodu, že od jara 2017 skončila intervence České národní banky. A protože je švýcarský frank jednou z globálních rezervních měn<sup>2</sup>, bylo by zajímavé vývoj tohoto kurzu analyzovat.

Data pro tuto práci byla čerpána ze stránek České národní banky <sup>3</sup> a analyzována v softwarovém programu MS Office Excel. Jelikož jde o denní kurz, ke studii je celkem  $n = 250$  pracovních dní neboli hodnot pozorování.

---

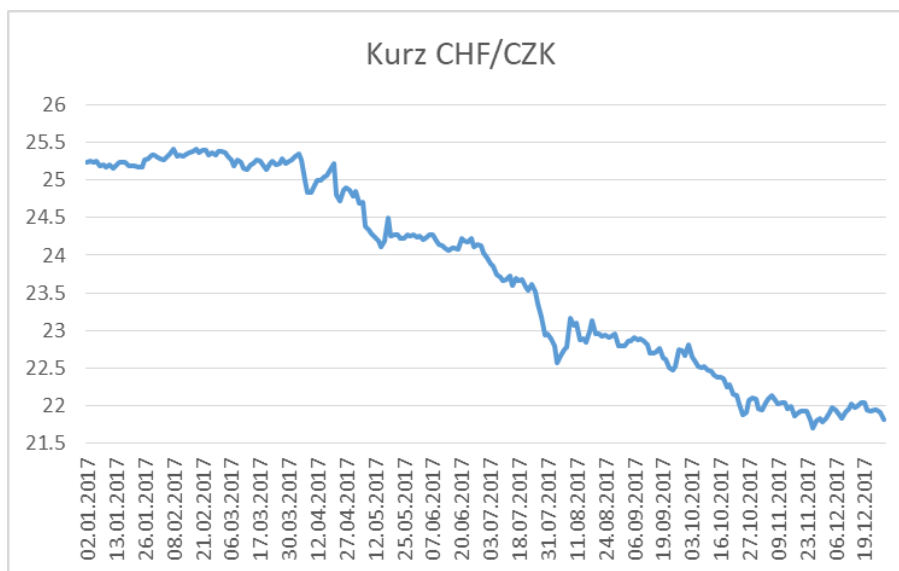
<sup>1</sup>CHF - švýcarský frank, CZK - česká koruna

<sup>2</sup>Rezervní měna je měna, která je ve větší míře uchovávána jako část zahraničních finančních rezerv. Nejzastoupenějšími rezervními měnami jsou americký dolar a euro.

<sup>3</sup>[https://www.cnb.cz/cs/financni\\_trhy/devizovy\\_trh/kurzy\\_devizoveho\\_trhu/vybrane\\_form.jsp](https://www.cnb.cz/cs/financni_trhy/devizovy_trh/kurzy_devizoveho_trhu/vybrane_form.jsp)

## 5.1. Základní charakteristika dat

Analýzu započnu vykreslením hodnot cen kurzu do grafu (obrázek 5.1). Denní kurz CHF/CZK je v případě tohoto spojnicového grafu vykreslen na vertikální ose. Na horizontální ose jsou vyznačeny vybrané dny z roku 2017. Jak je již na první pohled zřejmé, trend kurzu je jednoznačně klesající.



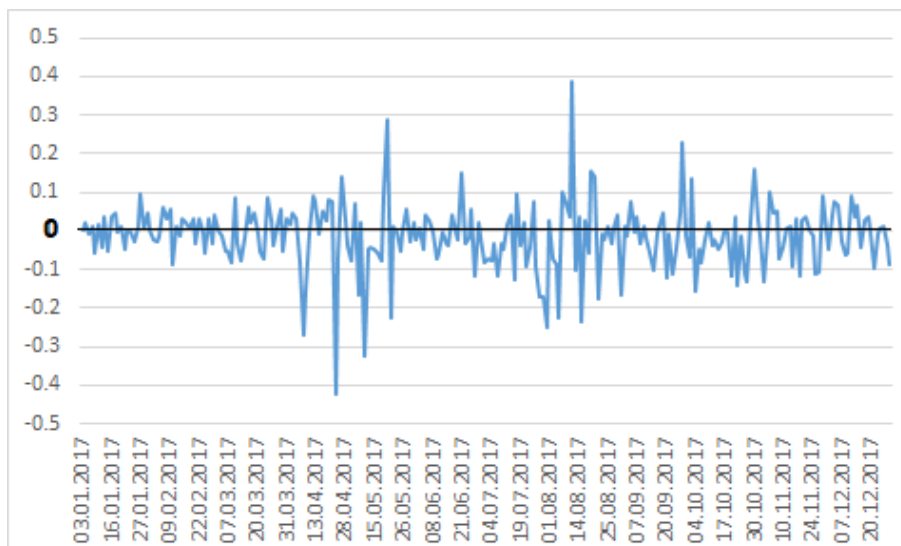
Obrázek 5.1: Časová řada denního kurzu CHF/CZK za rok 2017.

Při analýze časových řad je vhodné zjistit i jejich průměrné hodnoty. K tomu použiji vztahy pro popisné charakteristiky a míry dynamiky.

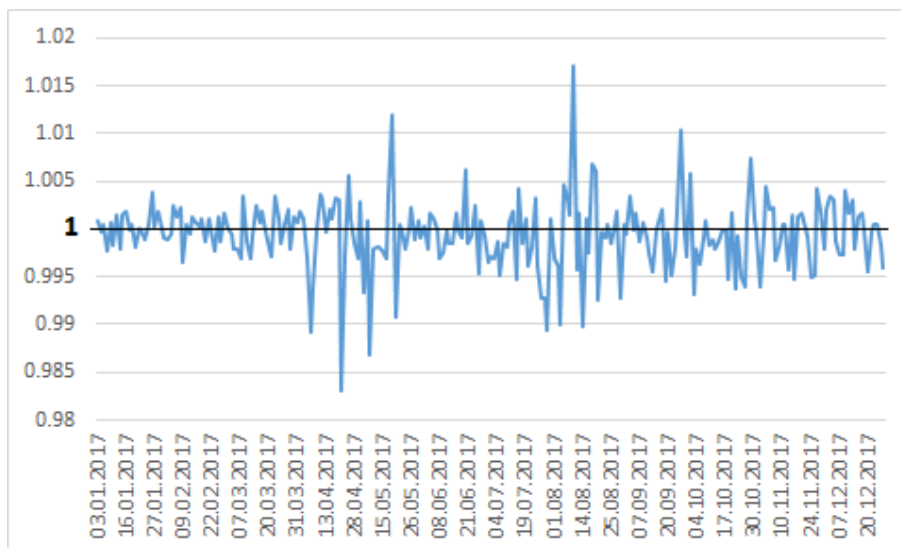
Pomocí prostého chronologického průměru zjistím, že průměrná hodnota denního kurzu CHF/CZK za rok 2017 je 23,73 korun českých.

Absolutní přírůstky, neboli první diference, jsem si vyjádřila graficky na obrázku 5.2. Průměrný absolutní přírůstek vychází  $-0,014$  korun českých. Toto číslo vyjadřuje, o kolik se průměrně změní hodnota kurzu za jeden den.

Koeficienty růstu jsem graficky znázornila na obrázku 5.3 a průměrný koeficient růstu, který vypočtu jako geometrický průměr jednotlivých koeficientů růstu, je roven hodnotě 0,999. Jelikož je hodnota menší než jedna, průměrná denní hodnota kurzu klesá.



Obrázek 5.2: Absolutní přírůstky kurzu CHF/CZK za rok 2017.



Obrázek 5.3: Koefficienty růstu kurzu CHF/CZK za rok 2017.

## 5.2. Vlastnosti finančních časových řad

V této části práce se pokusím otestovat, zdali platí jednotlivé předpoklady či vlastnosti, které jsou popsány v druhé kapitole.

### 5.2.1. Stacionarita

Co se týče stacionarity, již z obrázku 5.1, kde je graficky znázorněn denní kurz CHF/CZK, lze na první pohled vidět, že se hodnoty nevrací k žádné konstantě. Tím pádem typická vlastnost finančních řad – nestacionarita, platí i v tomto případě.

### 5.2.2. Volatilita

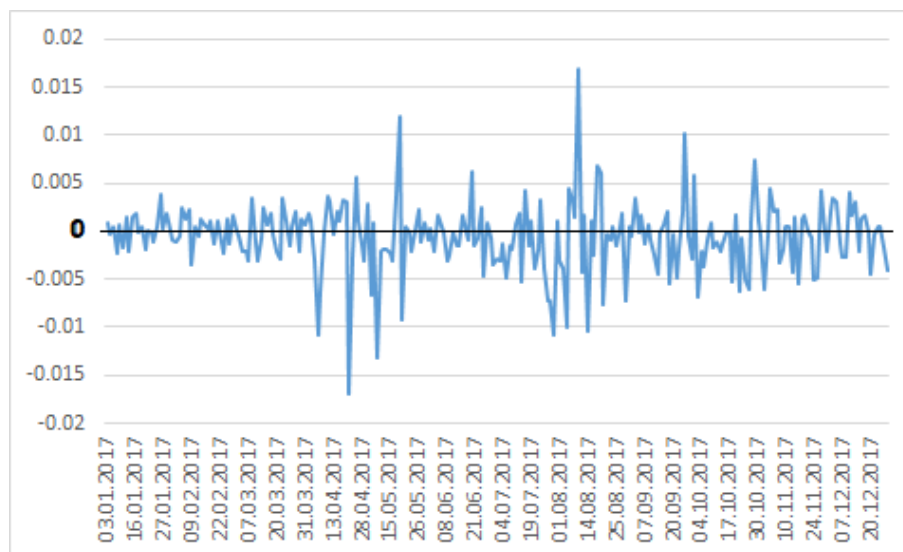
Při finančních časových řadách může volatilitu ovlivňovat několik faktorů. Jako jeden z nejdůležitějších faktorů je nesynchronní obchodování. Na finančním trhu se totiž obchoduje pouze v pracovních dnech. To má za následek shlukování volatility, z důvodu hromadění informací přes víkendy a následné obecné zvýšení volatility v pondělky. Také lze vyzorovat, že se i důležité summita konají o víkendu, aby se zamezilo okamžitým reakcím trhu.

Budoucí odhady chování volatility finančních časových řad jsou důležité především pro investory. Umožňuje jim posoudit míru rizikovosti dané investice.

Jedním ze způsobů zachycení volatility je její modelování pomocí speciálních modelů volatility. Pro tuto práci jsem však zvolila změřit volatilitu pomocí směrodatné odchylky neboli historickou volatilitu, přestože jsem si vědoma faktu, že ne vždy může být dostatečně vypovídající.

V tomto konkrétním případě se hodnota směrodatné odchylky logaritmu výnosů rovná číslu 0,004.

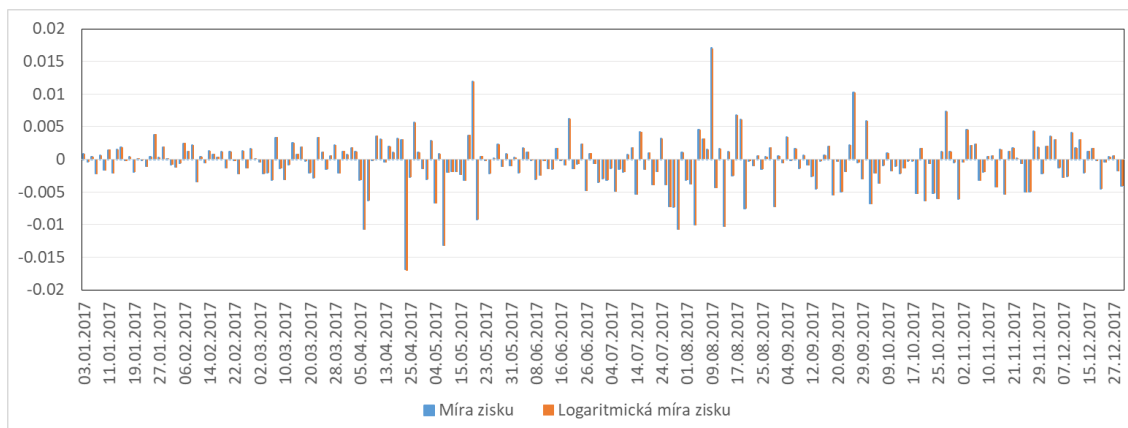
Na obrázku 5.4 lze vyzorovat typickou vlastnost – shlukování volatility.



Obrázek 5.4: Logaritmická míra zisku.

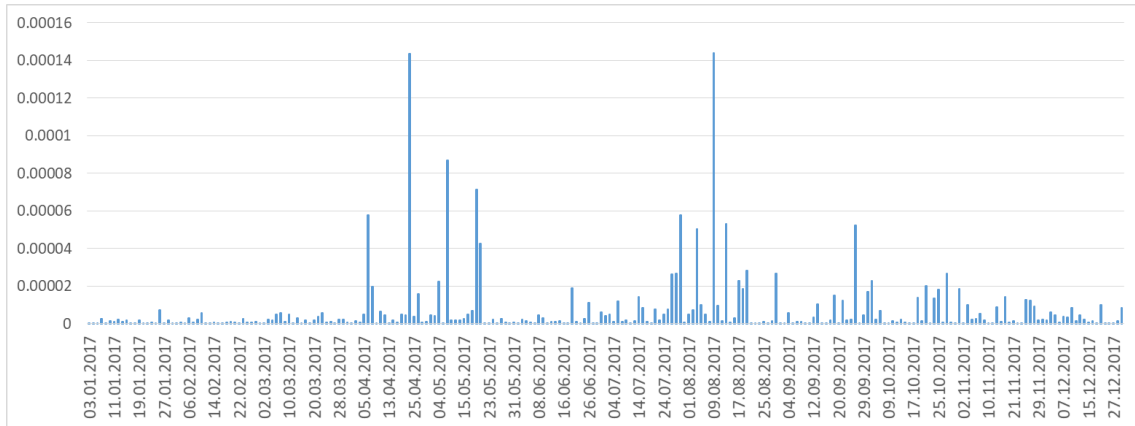
### 5.2.3. Leptokurtické rozdělení logaritmů výnosů

Jak jsem již zmínila, při malé výnosnosti se může považovat míra zisku i logaritmická míra zisku za stejné. Na obrázku 5.5a se opravdu tyto míry liší minimálně. Dá se říct, že v tomto rozlišení nejdou vůbec vidět. Z tohoto důvodu přikládám i graf prvních diferencí (obrázek 5.5b, abychom měli kvantitativní představu o jejich rozdílnosti. Dá se tedy konstatovat, že lze pracovat pouze s nezlogaritmovanou mírou zisku.



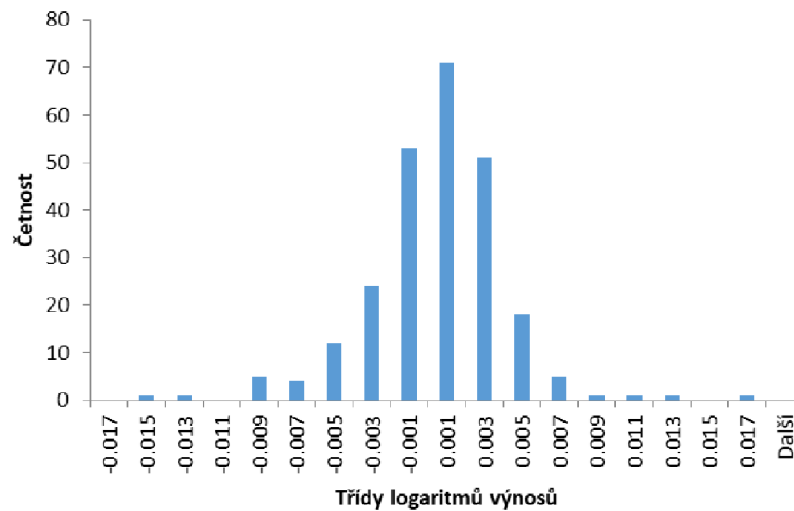
Obrázek 5.5a: Míra zisku a logaritmická míra zisku.





Obrázek 5.5b: Rozdíly mezi mírou zisku a logaritmickou mírou zisku.

Logaritmy výnosů nemají normální rozdělení. To lze ukázat výpočtem koeficientu šikmosti a koeficientu špičatosti nebo prostým pohledem na histogram (obrázek 5.6), který vyobrazuje četnost jednotlivých výnosů ve zvolených třídách.



Obrázek 5.6: Histogram logaritmických výnosů.

koeficient šikmosti $SK_r$	$-0,206$
koeficient špičatosti $K_r$	$4,477$

Tabulka 5.1: Koeficient šikmosti a špičatosti.

Záporný koeficient šikmosti  $SK_r$  značí zešikmení doleva, což znamená menší četnosti hodnot napravo od průměru, nežli nalevo. Rozdělení má tzv. *levý chvost*.

Koeficient špičatosti  $K_r$  je větší než tři, což se u časových řad měr zisků očekává. Znamená to tedy, že je rozdělení v porovnání s normálním rozdělením špičatější kolem středu a na chvostech má větší hustotu.

#### 5.2.4. Lineární nezávislost logaritmů výnosů

Lineární nezávislost lze ověřit pomocí Pearsonova korelačního koeficientu, přičemž je požadováno, aby jeho hodnota nabývala nuly. V tomto případě vyšel  $-0,03805$ , což je hodnota velmi blízká nule a mohu proto považovat logaritmy výnosů za lineárně nezávislé.

### 5.3. Dekompozice

#### 5.3.1. Analýza trendu

Analyzovaná data jsou na obrázku 5.1. Mým úkolem je zjistit jakousi základní tendenci vývoje, tudíž vybrat správnou trendovou funkci. Budu brát v úvahu trend lineární, kvadratický a případně exponenciální. Jednotlivé modely i s hodnocením jsou zaznamenány v tabulce 5.1.

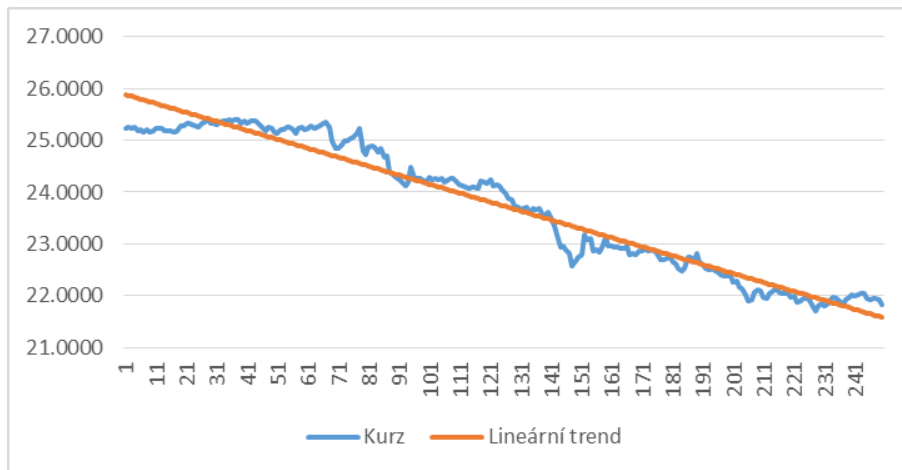
TREND	ROVNICE TRENDU	SSE	$R^2$
Lineární	$T_t = 25,902 - 0,017t$	19,317	0,953
Kvadratický	$T_t = 25,723 - 0,013t - 0,00001t^2$	–	–
Exponenciální	$T_t = 5,976 \cdot 0,999^t$	20,757	0,949

Tabulka 5.2: Modely trendů pro  $\forall t = 1, \dots, 250$ .

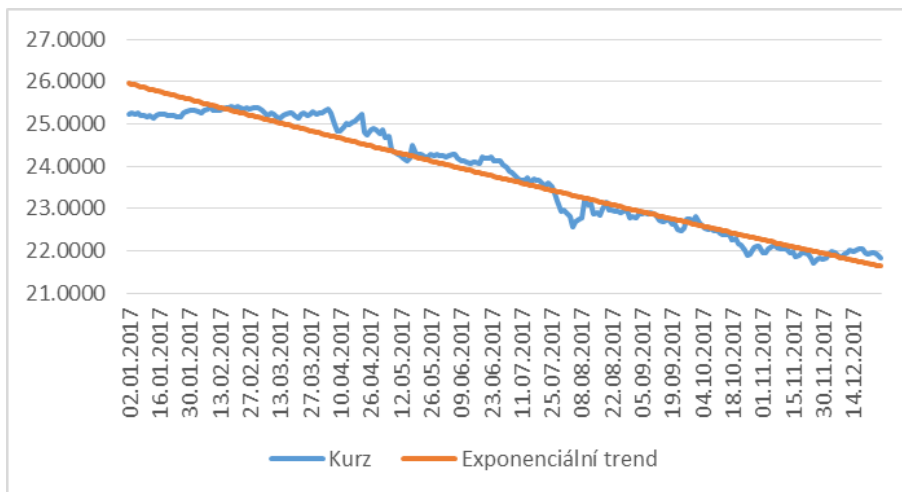
V případě kvadratického trendu je parametr  $\beta_2$  téměř nulový. Při provedení testu nulové hypotézy o nulovosti tohoto parametru jsem však hypotézu zamítla (p-value = 0,000017). Tato parabola proložila lépe začátek dat nežli konec. Nejspíš tomu tak je z důvodu, že první čtvrtina dat ještě podléhala intervenci ČNB a tudíž mohla zkreslit vhodnost trendu. K predikci tuto variantu nepovažuji za vhodnou. Dále tedy zvažuji jen trend lineární a exponenciální. Hodnoty  $SSE$  a  $R^2$  jsou v případě obou trendů velmi podobné. Pokud bych se však měla na základě těchto hodnot rozhodnout, kterým modelem budeme predikovat

budoucí hodnoty, zvolila bych trend lineární.

Při pohledu na oba grafy trendů (obrázek 5.7 a obrázek 5.8) je patrné, že exponenciální trend trochu lépe proložil koncové hodnoty, tudíž by mohl dát lepší a přesnější predikci. K analýze následujících složek tedy považuji exponenciální trend za vhodnější.



Obrázek 5.7: Lineární trend.

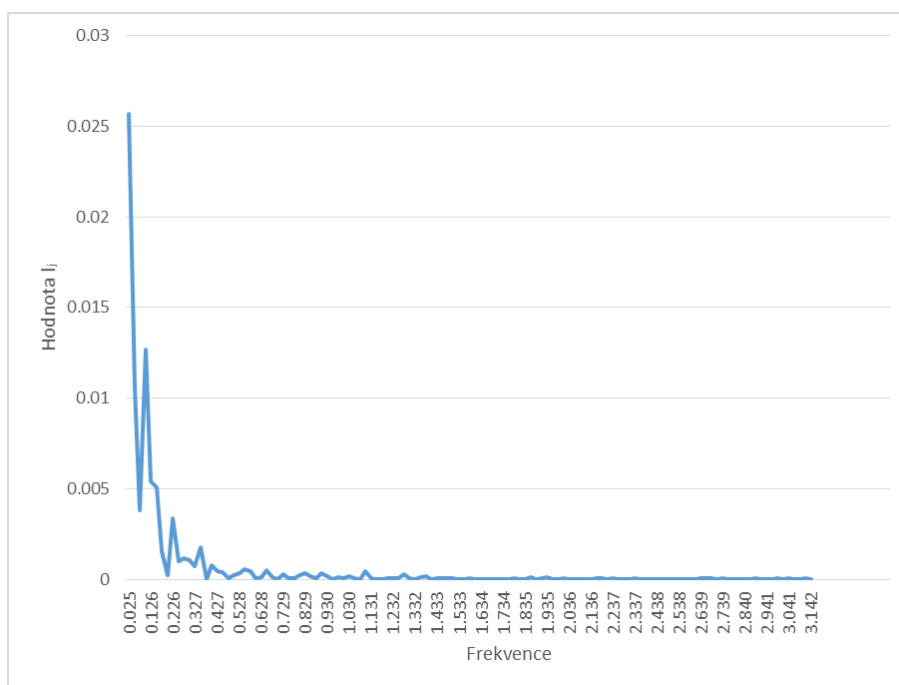


Obrázek 5.8: Exponenciální trend.

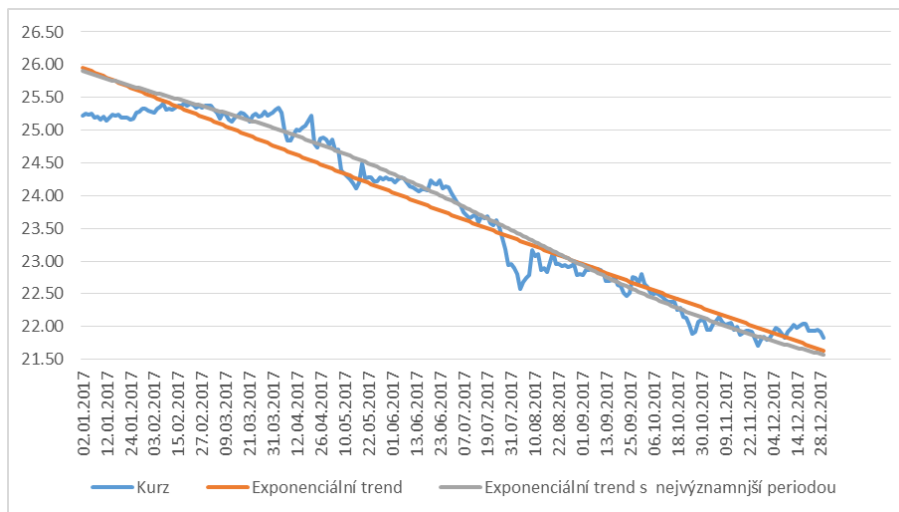
### 5.3.2. Analýza periodicity

Sezónní složku jsem se snažila najít pomocí periodogramu (obrázek 5.9). Při testování periodicity je nutné data detrendovat, a jelikož jsem zvolila jako základní tendenci trend exponenciální, je potřeba odečíst hodnoty tohoto modelu od reálných hodnot kurzu.

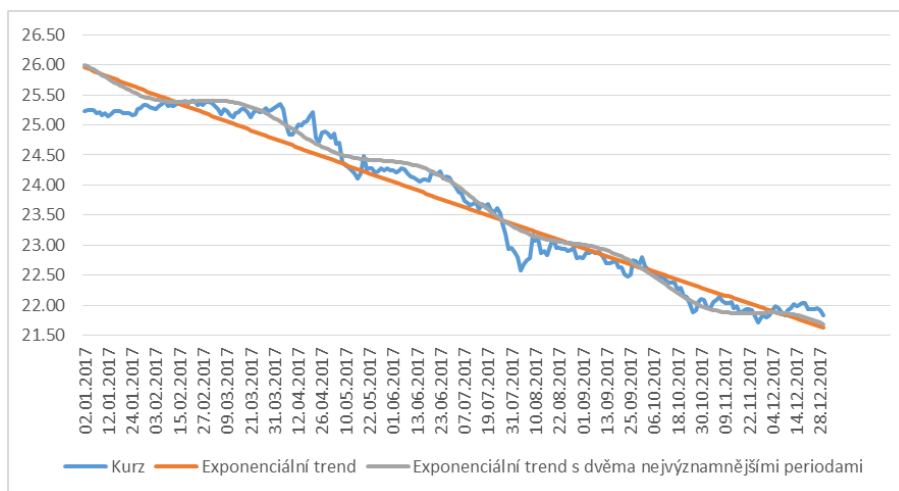
Fisherovým testem jsem následně zjistila, že je v datech celkem 13 významných period. Ke grafickému znázornění jsem vybrala pouze jednu, a potom dvě nejdůležitější periody. Nejvýznamnější perioda má frekvenci 0,0251, která by zřejmě v denních pozorováních představovala 250ti denní periodicitu. Druhá nejvýznamnější perioda by potom s frekvencí 0,1 vyjadřovala zhruba 63denní periodicitu.



Obrázek 5.9: Periodogram.



Obrázek 5.10: Exponenciální trend s nejvýznamnější periodou.



Obrázek 5.11: Exponenciální trend s dvěma nejvýznamnějšími periodami.

### 5.3.3. Analýza náhodné složky

Z obrázku 5.1 jde vidět, že kurz vykazuje jasný trend. Pro testování náhodnosti je nutné z modelu odstranit veškeré systematické složky jako je trend, sezónnost apod.

Abych vyloučila subjektivní názor na náhodnou složku, provedu test, který otestuje nulovou hypotézu  $H_0 : \epsilon_t \sim iid$ .

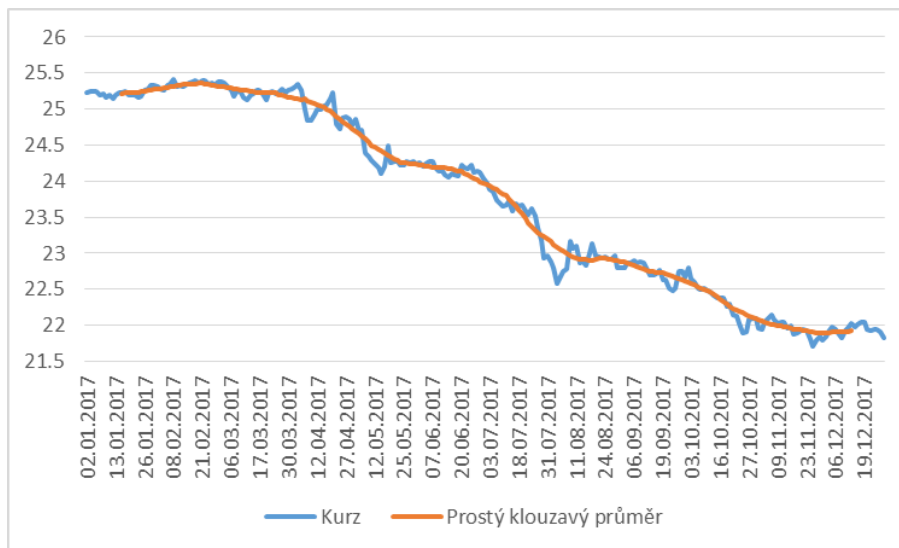
Jednotlivá data jsem si v prvním případě očistila jen od trendové složky a

vypočetla jednotlivá rezidua. Ty jsem upravila tím způsobem, že v případě dvou po sobě jdoucích stejných hodnot  $\epsilon_t$  jsem jednu z nich vynechala. Tato upravená data jsem následně analyzovala. Pomocí testu založeného na znaménkách diferencí jsem vypočetla, že statistika  $W$  nabývá hodnoty 3,826. Na hladině významnosti  $\alpha = 0,01$  ani  $\alpha = 0,05$  jsem nemohla nulovou hypotézu zamítnout.

V druhém případě jsem zkusila data očistit jak od trendové složky, tak od periodické. Celý postup jsem zopakovala. V tomto případě jsem opět nulovou hypotézu nemohla zamítnout na obou hladinách významnosti.

### 5.3.4. Klouzavé průměry

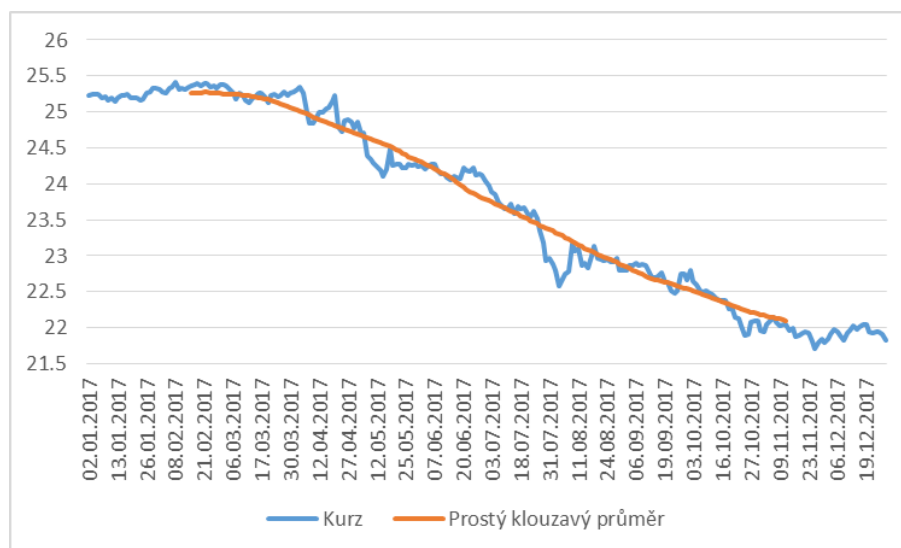
Průměrný měsíc má zhruba 21 pracovních dní. Z tohoto důvodu jsem se rozhodla využít necentrováné klouzavé průměry s délkou okna  $(2m + 1) = 21$ . Obrázek 5.12 zobrazuje právě tento model. Klouzavé průměry nevyhlazují začátek a konec dané řady, v závislosti na zvolení  $m$ . Predikci budoucích hodnot tak provedeme extrapolací koncové vyhlazené části vhodným polynomem.



Obrázek 5.12: Prostý necentrováný klouzavý průměr délky okna 21.

Pomocí klouzavých průměrů se převážně snažím vyhladit sezónní a periodické fluktuační. Z obrázku 5.12 je zřejmé, že v případě délky okna jednoho průměrného

měsíce se nejvýznamnější perioda nevyhladila. Proto zkusím volit délku okna o velikosti třech měsíců, tj  $(2m + 1) = 63$ . Jak jde vidět (obrázek 5.13), není sice vyhlazená větší část dat na počátku a na konci sledovaného období, data jsou však pěkně proložená a na první pohled nevykazují žádnou periodicitu. Při analýze periodické složky (podkapitola 5.3.2) pomocí periodogramu byla jedna z nejvýznamnějších period dlouhá zhruba 63 dní, mohla by to proto být správná délka okna.



Obrázek 5.13: Prostý necentrováný klouzavý průměr délky okna 63.

Tento model bude následně použit k predikci budoucích hodnot v podkapitole 5.4. Predikce.

### 5.3.5. Exponenciální vyrovnání

Exponenciální vyrovnání je další možná varianta způsobu vyhlazení časové řady a predikce. Pro tuto analýzu jsem zvolila dvojité exponenciální vyrovnání, které předpokládá, že je řada po částech lineární.

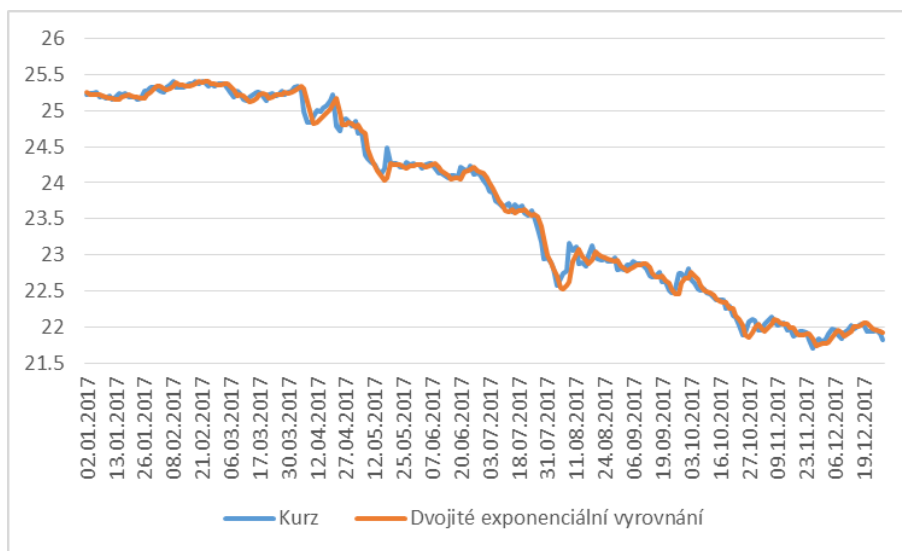
V první fázi je nutné zvolit vhodnou hodnotu vyrovnávací konstanty.

$\alpha$	SSE
0,1	3,06
0,2	2,68
0,3	2,44
0,4	2,31
0,5	2,26
0,6	2,32
0,7	2,52
0,8	2,9
0,9	4,3

Tabulka 5.3: Průběh SSE pro jednotlivé hodnoty vyrovnávací konstanty  $\alpha$ .

Dle doporučení by měla být optimální varianta v intervalu  $\alpha \in \langle 0,7; 1 \rangle$ . Podle tabulky 5.3 poskytuje optimální výsledky volba  $\alpha = 0,5$ , což by mohlo znamenat, že tato metoda není příliš vhodná. Při jednoduchém exponenciálním vyrovnání jsem však narazila na stejný problém.

Pokud se budu držet doporučení  $\alpha \in \langle 0,7; 1 \rangle$  a vyberu nejlepší variantu, bude to  $\alpha = 0,7$ . Graficky tedy znázorním časovou řadu kurzu s dvojitým exponenciálním vyrovnáním a vyrovnávací konstantou  $\alpha = 0,7$ .



Obrázek 5.14: Dvojitě exponenciální vyrovnání.

Předpovědi budoucích hodnot podle dvojitého exponenciálního vyrovnání budou taktéž vypočteny v podkapitole 5.4. Predikce.



## 5.4. Predikce

Predikci jsem dělala na měsíc leden roku 2018.

Jako první se pokusím zkontrolovat volbu trendu. Při výběru jsem zvolila trend exponenciální na základě mého subjektivního názoru, nikoli podle výpočtů  $SSE$  nebo  $R^2$ . Proto zkusím porovnat predikce trendu exponenciálního s predikcí trendem lineárním, abych zjistila, jestli jsem opravdu zvolila lepší variantu.

K dalším predikcím využiji adaptivní přístupy k modelování trendu, jako jsou klouzavé průměry a exponenciální vyrovnání.

Poslední variantou bude predikce pomocí martingalu.

Poznámka.  $SSE$  chápeme jako součet čtvercových chyb jednotlivých předpovědí za celý měsíc.

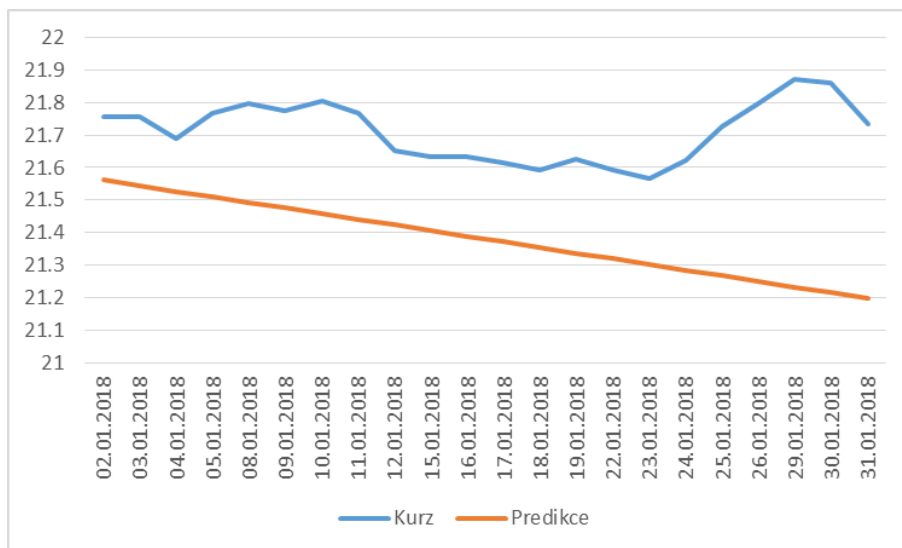
### 5.4.1. Model 1 – lineární trend

V prvním případě jsem predikovala budoucí hodnoty lineárním trendem. Predikční model má tvar

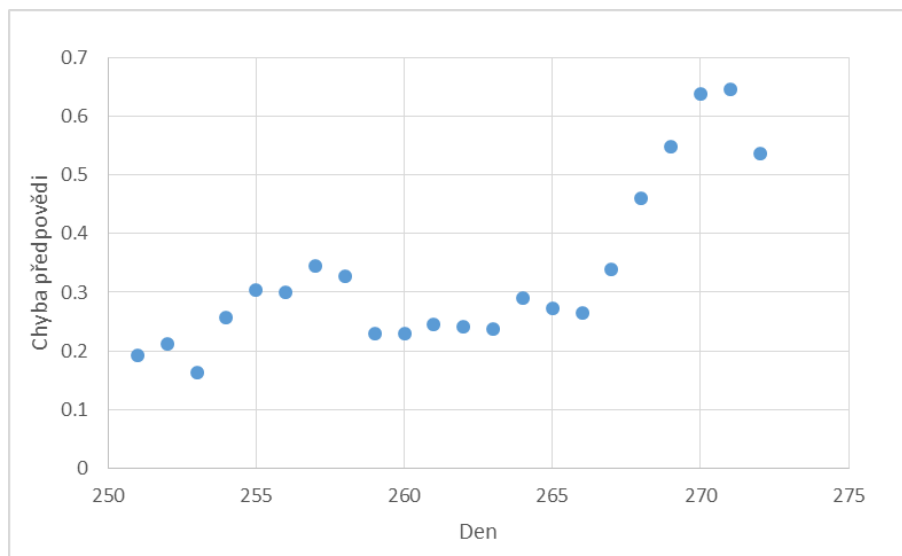
$$\hat{y}_T = 25,902 - 0,017T, \quad \forall T = 251, \dots, 272.$$

Součet chyb předpovědí je roven hodnotě 2,829.

Po grafickém znázornění chyb předpovědí si lze všimnout, že model mnohem lépe predikuje hodnoty prvních zhruba patnácti dní, nežli dní pozdějších. Tento model není příliš přesný už jen z důvodu, že průměrná hodnota předpovědní chyby je rovna číslu 0,331 a její směrodatná odchylka hodnotě 0,141.



Obrázek 5.15: Predikce modelem 1.



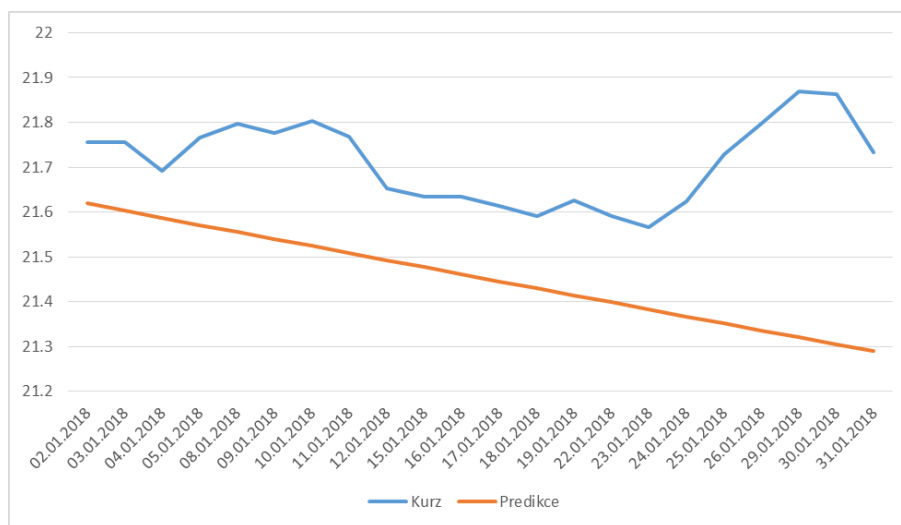
Obrázek 5.16: Graf chyb předpovědi modelu 1.

## 5.4.2. Model 2 – exponenciální trend

Druhým případem je model s exponenciálním trendem tvaru

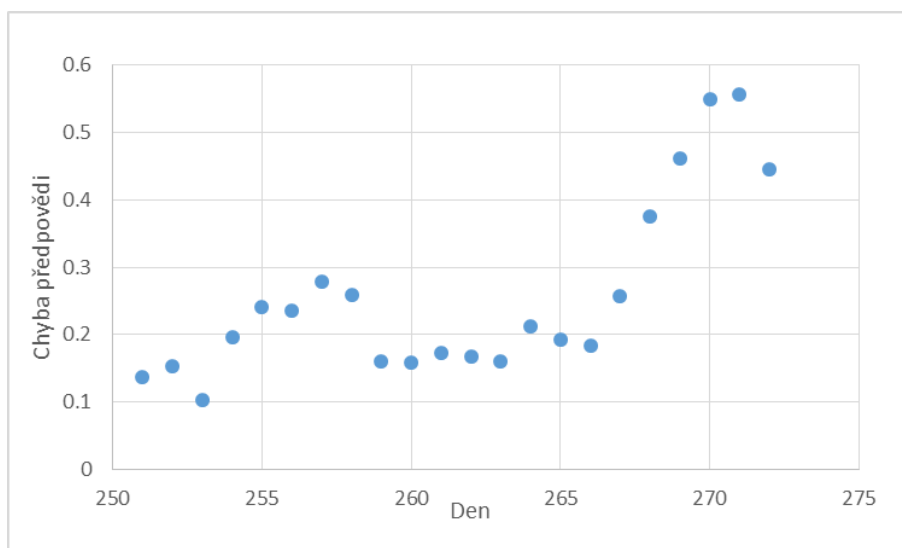
$$\hat{y}_T = 5,976 \cdot 0,999^T, \quad \forall T = 251, \dots, 272.$$

Součet čtvercových chyb je v tomto případě roven hodnotě 1,832. Tímto se potvrdilo, že model s exponenciálním trendem je vhodnější.



Obrázek 5.17: Predikce modelem 2.

I v tomto modelu je přesněji predikováno zhruba prvních patnáct dní, jak je možno vidět z obrázku 5.18. Průměrná hodnota předpovědní chyby je 0,257. Směrodatná odchylka nabývá hodnoty 0,137.



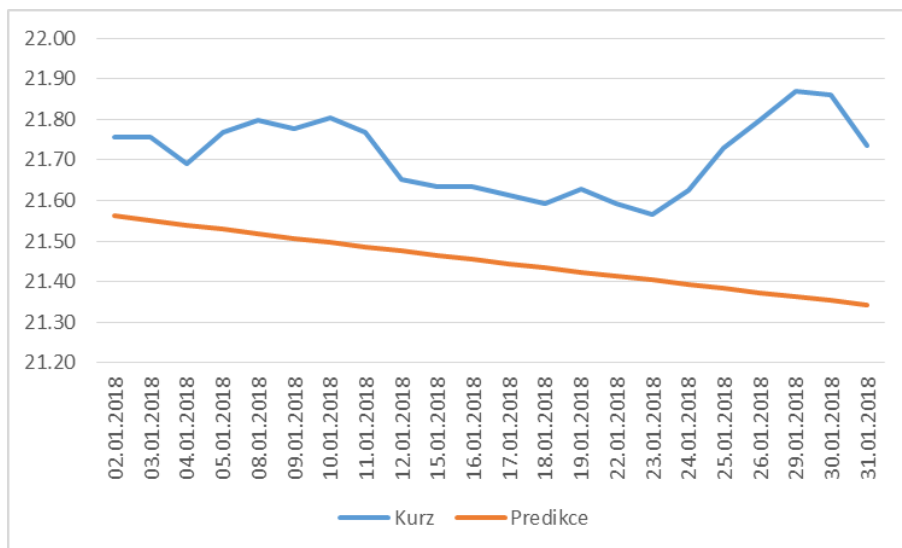
Obrázek 5.18: Graf chyb předpovědí modelu 2.

### 5.4.3. Model 3 – exponenciální trend s periodou

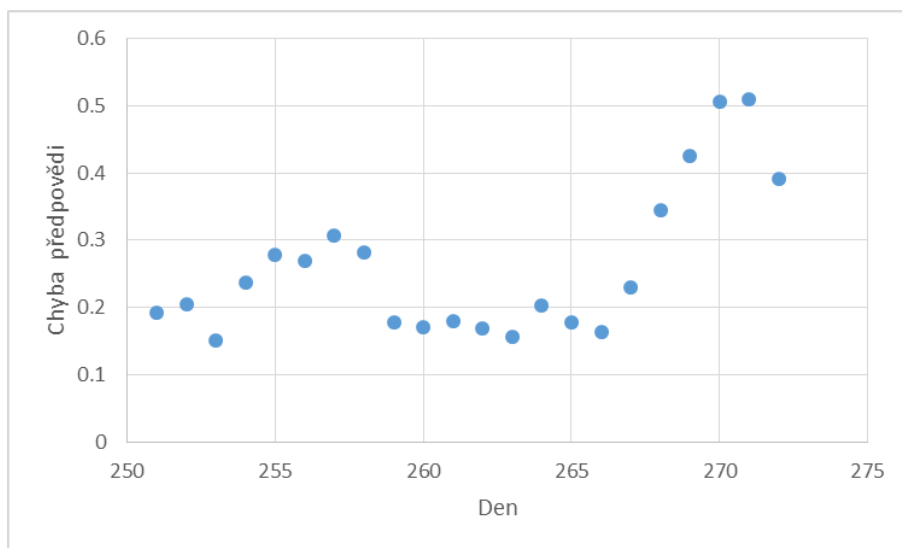
Jelikož jsem se ujistila ve správném výběru trendu, tzn. exponenciálního, zkusím do modelu zahrnout i nejvýznamnější periodu. Model bude ve tvaru

$$\hat{y}_T = 5,976 \cdot 0,999^T - 0,138 \cdot \cos\left(\frac{\pi}{125}T\right) + 0,18 \cdot \sin\left(\frac{\pi}{125}T\right), \quad \forall T = 251, \dots, 272.$$

V tomto modelu se nepatrně snížil součet čtvercových chyb na hodnotu 1,75. Průměrná hodnota předpovědní chyby je 0,26 a její směrodatná odchylka 0,11.



Obrázek 5.19: Predikce modelem 3.



Obrázek 5.20: Graf chyb předpovědí modelu 3.

#### 5.4.4. Model 4 – klouzavé průměry

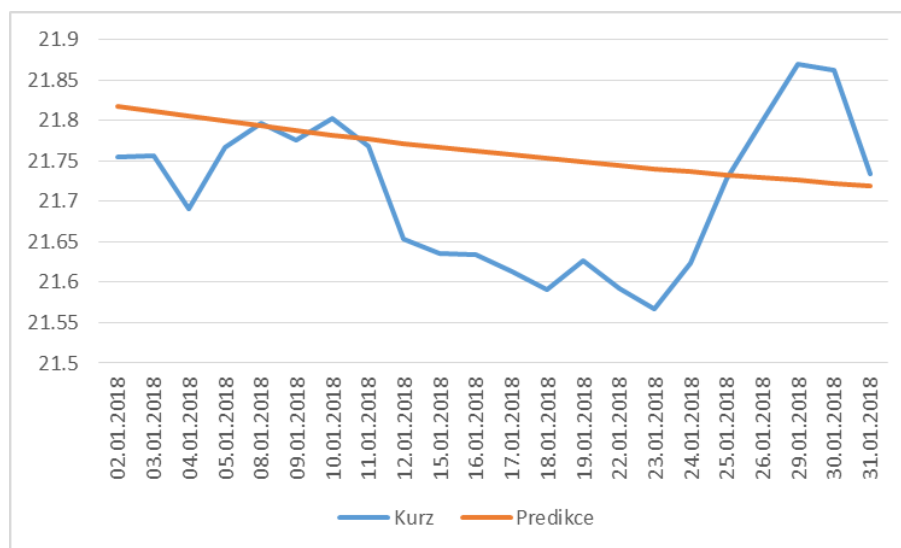
Další varianta, jak lze predikovat budoucí hodnoty, je pomocí klouzavých průměrů. V podkapitole 5.3.4 jsem vyhodnotila, že délka okna  $(2m + 1) = 63$  je ideální pro predikce, jelikož nevykazuje známky periodicity. Zkusím proložit poslední vyrovnané okno, tj. tři měsíce, trendem lineárním a kvadratickým. Při první možnosti – trend lineární se součet čtvercových chyb SSE blížil hodnotě 5, což je

mnohem horší než v předchozích modelech. V případě druhém – kvadratický trend (obrázek 5.21) se dá predikce považovat za velmi přesnou. Může nás zmást predikční rovnice tvaru

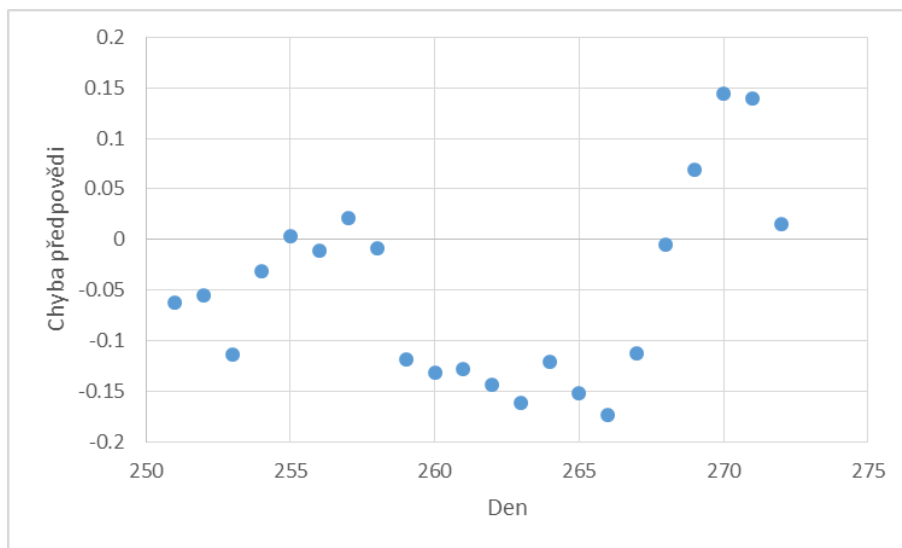
$$\hat{y}_T = 28,279 - 0,0452T + 0,00007T^2,$$

kdy parametr  $\beta_2$  nabývá velmi nízké hodnoty. Pro predikci se však tato varianta jeví jako jedna z nejlepších. Součet čtvercových chyb je roven hodnotě 0,242.

Na obrázku 5.21 jde vidět, že průměrná hodnota chyby předpovědi je  $-0,052$  a směrodatná odchylka je 0,093. Znamená to tedy, že tento model je z předchozích modelů nejpřesnější.



Obrázek 5.21: Predikce modelem 4.



Obrázek 5.22: Graf chyb předpovědi modelu 4.

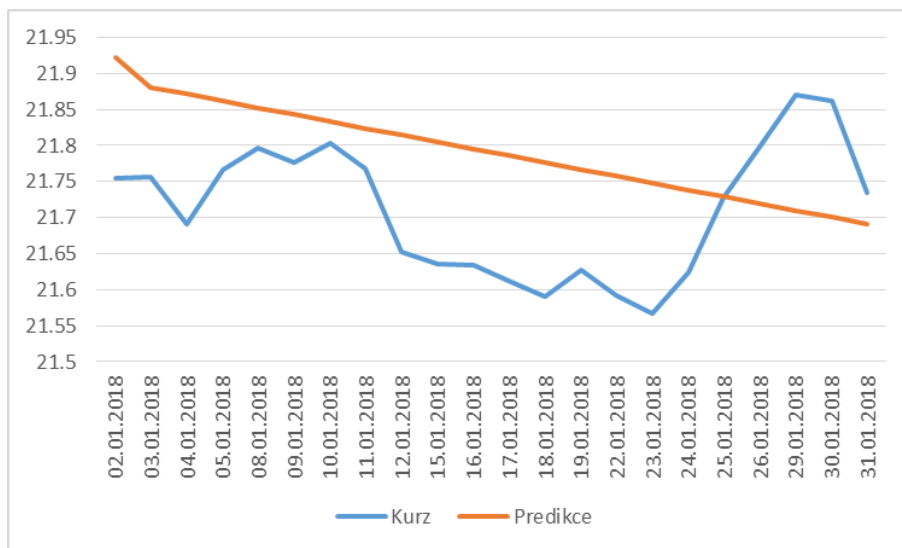
#### 5.4.5. Model 5 – exponenciální vyrovnání

V tomto případě využijí k predikci budoucích hodnot kurzu dvojitě exponenciální vyrovnání. Zvolím model s vyrovnávací konstantou  $\alpha = 0,7$ .

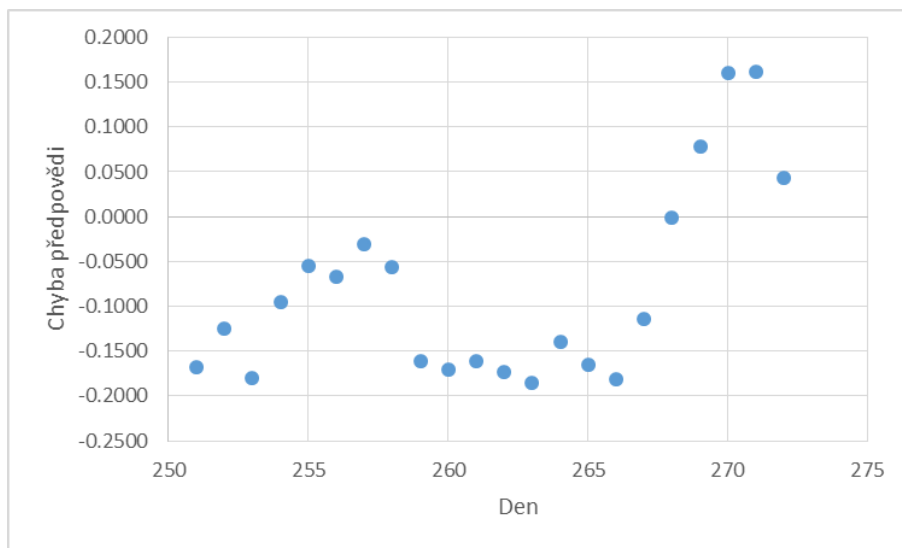
Postupně provedu predikce v čase  $t = 250$  na čas  $t + \tau$ , kde  $\tau = 1$ . Získám tak k dispozici hodnotu

$$\hat{y}_{251}(250) = \hat{\beta}_0(250) + \hat{\beta}_1(250) \cdot \tau = 21,93 - 0,007 \cdot 1 = 21.92,$$

díky které můžu vypočítat statistiky  $S_{25}$  a  $S_{25}^{[2]}$ , dále odhady  $\hat{\beta}_0$  a  $\hat{\beta}_1$ . Následně můžeme pokračovat a počítat predikci v čase  $t = 251$  na čas  $t + \tau$ , kde  $\tau = 1$  a tento postup opakovat až do  $\hat{y}_{272}(271)$ .



Obrázek 5.23: Predikce modelem 5.



Obrázek 5.24: Graf chyb předpovědí modelu 5.

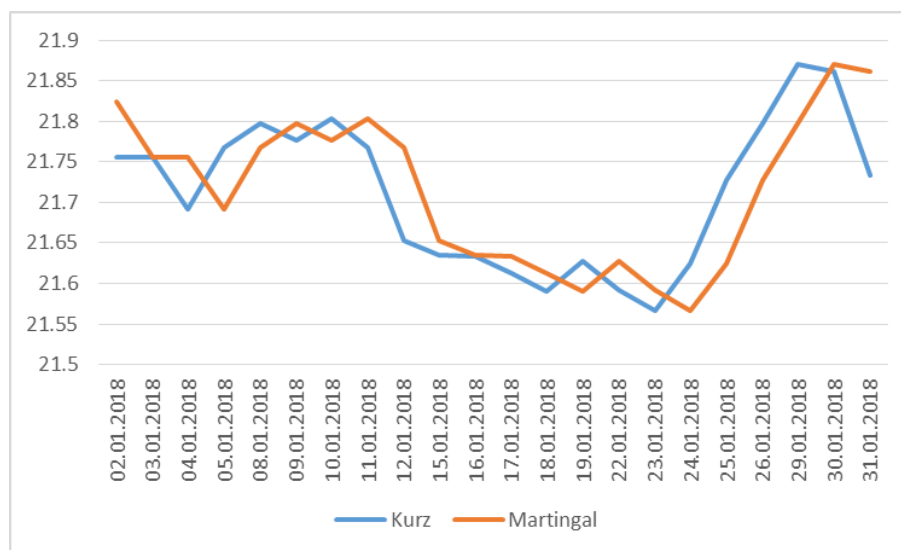
Střední čtvercová chyba predikce je rovna hodnotě 0,39. Průměrná hodnota chyby předpovědi je  $-0,081$  a směrodatná odchylka 0,11.

Poznámka. Když se vrátím k variantě  $\alpha = 0,5$ , která vyšla jako optimální, střední čtvercová chyba vyšla jen nepatrně menší.



### 5.4.6. Model 6 – martingal

K modelu 6 využijí martingal. Predikci dat pomocí martingalu lze chápat jako predikci zítřejší hodnoty tou dnešní. Je nutné si uvědomit, že tento model jsme mohli zkonstruovat zpětně až ve chvíli, kdy nám byly známy reálné hodnoty z tohoto období. Po shlédnutí obrázku 5.25 je možné vidět, že křivka martingalu vypadá naprosto stejně, pouze je posunutá o jeden den nazpátek. Tato predikce je samozřejmě jedna z nejpřesnějších, což dokazuje i SSE, který se rovná pouhých 0,076. Martingal také neříká žádné informace o základní tendenci vývoje či jiných systematických složkách analyzované řady. Nelze tedy říct, jak se bude kurz vyvíjet do budoucna, ani popsat jeho chování v minulém období.



Obrázek 5.25: Martingal.

Martingal se dá použít ve fundamentální analýze. Ta předpokládá, že vnitřní hodnoty, například akcií, nezávisí na jejich historických datech, pouze na rozdílu mezi vnitřní a tržní hodnotou. To znamená, že analytici této metody věří, že finanční data nepodléhají žádnému trendu a budoucí hodnoty tudíž nelze predikovat pomocí technické analýzy. S využitím martingalu ji lze jen s určitou pravděpodobností odhadnout při modelování cen akcií.

## 5.5. Hodnocení predikčních modelů

Tabulka 5.4 udává souhrnné informace o přesnosti predikcí dle jednotlivých modelů. V prvním sloupečku jsou jednotlivé výše popsané modely. Dále je zde součet čtvercových chyb predikcí a následně průměrná hodnota předpovědní chyby a směrodatná odchylka předpovědní chyby.

Poznámka. ch.p. = chyba předpovědi.

Model	SSE	průměrná ch.p.	směrodatná odchylka ch.p.
1	2,83	0,33	0,14
2	1,83	0,26	0,14
3	1,75	0,26	0,11
4	0,242	-0,052	0,093
5	0,39	-0,081	0,11
6	0,076	-0,004	0,06

Tabulka 5.4: Srovnání predikčních modelů.

### Hodnocení dle SSE

Nejllepší výsledek SSE má jednoznačně martingal. Jak jsem již napsala výše, je zkonstruován zpětně za předpokladu, že již známe data predikovaného období. Do celkového hodnocení ho proto nezahrnuji.

Jednoznačně nej přesnější model vykazující hlavní tendenci kurzu, je model zkonstruovaný pomocí klouzavých průměrů historických dat a následným proložením posledního vyrovnaného okna trendem kvadratickým. Mám na mysli model 4. Jde tedy vidět, že při správném zvolení délky vyrovnávacího okna, v tomto případě  $(2m + 1) = 63$ , lze poměrně dobře vyhladit časovou řadu a odstanit tím periodické fluktuace. Tato predikce je tudíž velmi zdařilá.

Přestože jsem měla problém při volbě vyrovnávací konstanta  $\alpha$ , která nejspíše indikovala jakousi chybu, predikce pomocí exponenciálního vyrovnání se dá považovat také za velmi přesnou. Lze tedy říct, že pomocí adaptivních přístupů k modelování trendu jsem dostala nejlepší predikce.

První dva modely jsou velmi jednoduché. Jde tedy o případy, kdy vycházím pouze z trendových rovnic, vypočítaných z historických dat. Pokud by bylo cílem

získat jen hrubý odhad budoucích hodnot, tyto modely by mohly posloužit vcelku dobře.

O něco málo lepších výsledků, než u prvních dvou modelů, jsem dostala predikcí modelem 3. Zde byl použit trend exponenciální s jednou nejvýznamnější periodou. Pro zajímavost jsem si sestrojila ještě jeden model, který byl stejný, pouze v něm byla zahrnuta i druhá nejvýznamnější perioda. Součet čtvercových chyb tohoto doplňujícího modelu byl 2,66.

Jde tedy vidět, že neplatí úměra – čím přesnější model chování historických dat, tím přesnější predikce. Historická data nám sice tento model proložil hezky, u predikce však selhal.

### **Hodnocení dle jednotlivých chyb předpovědi**

Grafy chyb předpovědí prvních dvou modelů se mohou jevit dost stejně. Model 2 má však jednotlivé chyby pro jednotlivé časy menší zhruba o 30%, tudíž je o něco přesnější. V obou případech je prvních asi 15 dní predikováno s větší přesností než dny další. Průměrná hodnota předpovědní chyby je však v obou případech dost velká.

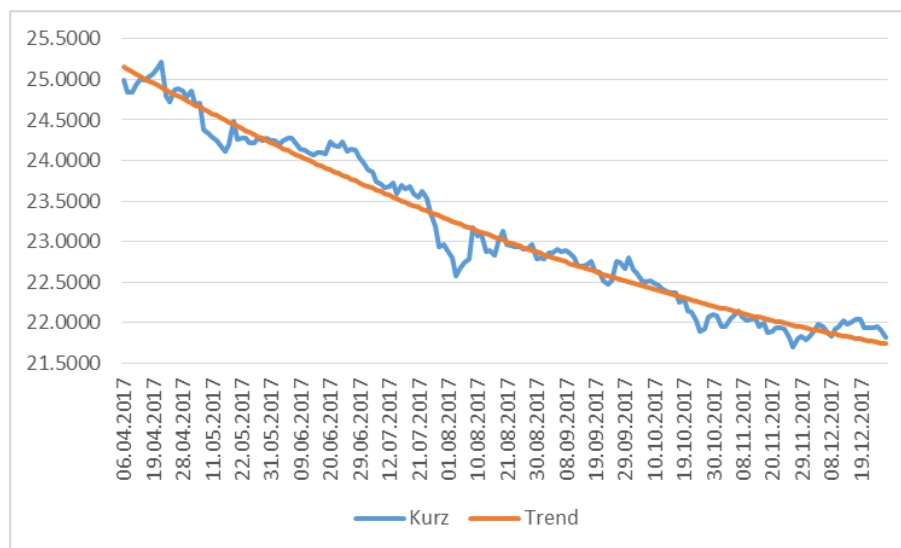
Přesnost predikce modelu 3 je bohužel hodně podobná jednoduchým modelům typu 1 a 2.

Model 4 jsem zvolila za nejlepší. Příslušný graf chyb předpovědí ukazuje, že jednotlivé hodnoty ve všech okamžicích téměř nepřesáhnou hodnotu 0,15. Můžu tedy říct, že tento model predikoval opravdu s nízkou chybovostí celý měsíc. Průměrná hodnota předpovědní chyby je v případě pátého modelu jen málo vyšší než při modelu 4, lze tedy tyto dva modely považovat za nejpřesnější.

## 5.6. Analýza kuzru CHF/CZK po skončení intervence ČNB

Předchozí analýzu jsem dělala z historických dat celého roku 2017. Jak jsem již zmínila, od jara 2017 skončila intervence ČNB, konkrétně tedy k 6.4.2017. Když se podíváme na obrázek 5.1, kde byly ukázána všechna data, jde vidět stálost kurzu v první čtvrtině roku. Zkusím tedy celou analýzu provést znovu, pouze s daty po skončení intervence.

Při analýze trendu jsem zjistila, že pro tuto řadu je nejvhodnější trend kvadratický, který proložil historická data lépe zhruba o 35%, nežli model lineární a exponenciální.



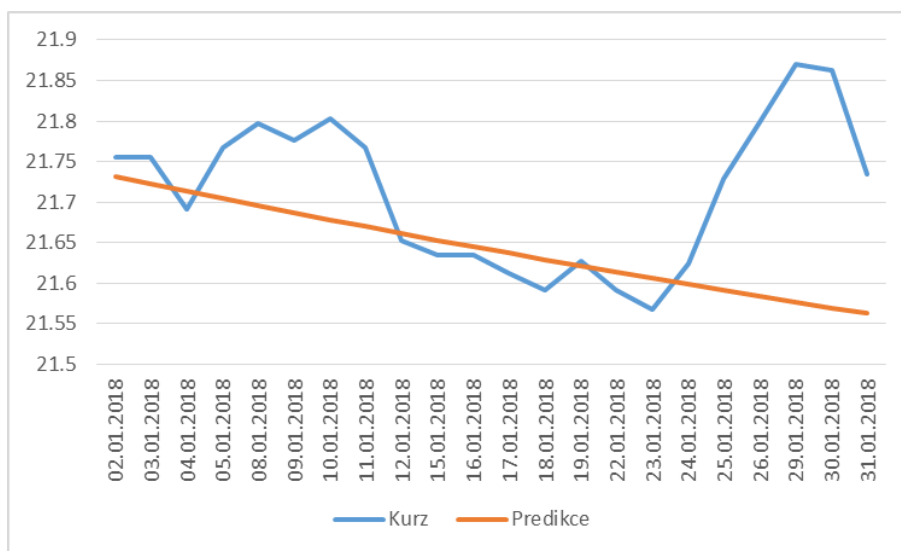
Obrázek 5.26: Kvadratický trend.

Fisherovým testem periodicity jsem určila 5 významných period s frekvencemi (1)0, 104, (2)0, 242, (3)0, 345, (4)0, 311, (5)0, 138.

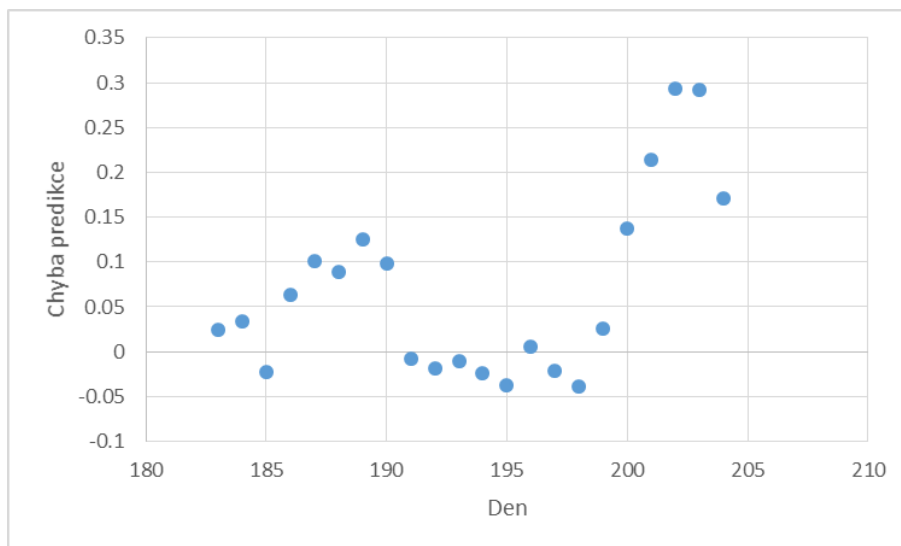
Při analýze náhodné složky jsem nemohla zamítnout nulovou hypotézu  $H_0 = \epsilon_t \sim iid$  na hladině významnosti  $\alpha = 0,01$  i  $\alpha = 0,05$ . Znamená to tedy, že v reziduální složce nejsou známky systematického chování.

K predikci jsem zkusila použít kvadratický trend. Součet predikčních chyb

vyšel 0,32, průměrná predikční chyba je rovna hodnotě 0,068 a směrodatná odchylka 0,099. V případě, kdy jsem do predikčního modelu ke kvadratickému trendu přidala nejprve jednu nejvýznamnější, a pak i ostatní výše uvedené periody, součet predikčních chyb ani v jednom z pěti případů nepřekročil hodnotu 3,9. Takovou chybovost považuji za velmi špatnou, proto tyto modely ani graficky neukazují.



Obrázek 5.27: Predikce kvadratickým trendem.



Obrázek 5.28: Chyby předpovědí.

Je nutné si uvědomit, že v předchozích modelech (5.4. Predikce) jsme pracovali s časovou řadou denních hodnot kurzu bez jakéhokoli ohledu na onu intervenci ČNB. Je proto zřejmé, že první čtvrtina dat se téměř neodchylovala od hodnoty 25,2 a tato skutečnost mohla dost ovlivnit predikční model. Proto jsem se v této části práce rozhodla tato data vyřadit a analyzovat pouze data, kterých se již intervence netýkala.

Pokud bych měla porovnat jednoduché modely, tj. tvořeny pouze trendem, v případě zohlednění ekonomické situace jsem predikovala mnohem přesněji. Je tudíž důležité brát ohled na chování trhu.

# Závěr

V první části této bakalářské práce jsem popsala teorii časových řad, přístupy k analýze a následné možnosti predikce. Také jsem uvedla druhy časových řad a popsala vlastnosti, které dělají z časové řady řadu finanční. Okrajově jsem zmínila i možnosti predikce pomocí martingalu.

Cílem této práce byla analýza konkrétní finanční časové řady, zkonstruování možných predikčních modelů a následné hodnocení predikcí. Ke studii jsem si vybrala finanční časovou řadu denního kurzu CHF/CZK v období od 1.1.2017 do 31.12.2017. Následná predikce budoucích hodnot kurzu byla prováděna na měsíc leden roku 2018.

Při analýze chování tohoto kurzu jsem se nejprve věnovala jednotlivým složkám dekompozice, následně jsem vytvořila šest predikčních modelů, kterými jsem předpovídala budoucí hodnoty. Porovnání přesnosti jednotlivých modelů jsem shrnula v tabulce 5.4, pod kterou jsem jednotlivé modely ohodnotila detailněji. Závěrem bych jen zdůraznila, že jelikož jsou adaptivní přístupy schopny pracovat s měnícím se charakterem trendu, jejich využití pro predikci se jeví jako nejpřesnější.

Nakonec jsem zkusila predikovat budoucí hodnoty za předpokladu znalosti ekonomické situace. Analyzovala jsem tedy historická data od ukončení intervence ČNB a na základě jejich chování provedla predikci. Jak jsem předpokládala, i jednoduché modely tvořící pouze trendovou složku predikovaly přesněji, než v případě analýzy celého roku, neboť první čtvrtina roku modely nezakreslila.

Závěrem lze říct, že k hrubé představě o chování kurzu nám dobře posloužily jednoduché modely, jejichž základem je pouze trend. Pokud zvolená časová řada

vykazuje jasnou periodicitu, je nutné dobře zvážit, jestli je vhodné ji do predikčního modelu zahrnovat.

Finanční časové řady jsou v některých případech tvořeny pouze bílým šumem. Z tohoto důvodu by mohlo být zajímavé srovnání klasických modelů, například s modely Box-Jenkinsonovy metodologie, které jsou založeny na analýze náhodné složky. Toto srovnání by mohlo být náplní mé další práce v navazujícím studiu.



# Literatura

- [1] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL, Praha, 1986.
- [2] Artl, J., Artlová, M.: *Finanční časové řady*. Grada, Praha, 2003.
- [3] Hindls, R., Kaňoková, J., Novák, I.: *Metody statistické analýzy pro ekonomy*. Management Press, Praha, 1997.
- [4] Prášková, Z., Lachout, P.: *Základy náhodných procesů*. Karolinum, Praha, 2005.
- [5] Cipra, T.: *Finanční ekonometrie*. Ekopress, Praha, 2008.
- [6] Janík, V.: *Diferenciální počet I*. Academia, Praha, 1974.
- [7] Hušek, R.: *Základy ekonomické analýzy II*. Vysoká škola ekonomická v Praze, 1998.
- [8] Artl, J., Artlová, M.: *Ekonomické časové řady*. Grada, Praha, 2007.
- [9] Cipra, T.: *Kapitálová přiměřenost ve financích a solventnost v pojišťovnictví*. Ekopress, Praha, 2002.
- [10] Artl, J., Artlová, M., Rublíková, E.: *Analýza ekonomických časových řad s příklady*. VŠE, Praha, 2002.  
Dostupné z: <http://nb.vse.cz/artlova/vyuka/crsbir02.pdf>.
- [11] Anděl, J.: *Matematická statistika* SNTL, Praha, 1978.
- [12] Musil, K.: *Sezónnost a periodičita v časových řadách*. UK v Praze, Praha, 2011.  
Dostupné z: [http://dspace.cuni.cz/bitstream/handle/20.500.11956/37867/BPTX\\_2010\\_1\\_\\_0\\_259343\\_0\\_91600.pdf](http://dspace.cuni.cz/bitstream/handle/20.500.11956/37867/BPTX_2010_1__0_259343_0_91600.pdf).
- [13] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky*. Olomouc, 2015.

- [14] Tsay, Ruey, S.: *Analysis of Financial Time Series, second edition*. John Wiley & Sons, INC., New York, 2002. Dostupné z: <http://cfa.goldenglobal.org.cn/uploadfile/appendfile/%E8%B5%84%E6%96%99%E4%B8%8B%E8%BD%BD/CFA%E5%AD%A6%E4%B9%A0%E8%B5%84%E6%96%99/Analysis%20of%20Financial%20Time%20Series%202nd%20Edition.pdf>.