

# Analýza a modelování burzovních dat

Bc. Roman Směták

---

Diplomová práce  
2019



Univerzita Tomáše Bati ve Zlíně  
Fakulta aplikované informatiky

---

Univerzita Tomáše Bati ve Zlíně  
Fakulta aplikované informatiky  
akademický rok: 2018/2019

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Roman Směták**  
Osobní číslo: **A15443**  
Studijní program: **N3902 Inženýrská informatika**  
Studijní obor: **Informační technologie**  
Forma studia: **kombinovaná**

Téma práce: **Analýza a modelování burzovních dat**  
Téma anglicky: **Analysis and Modelling Stock Exchange Data**

Zásady pro vypracování:

1. Získejte data časových řad na základě vývoje vybraných akciových titulů popřípadě akciových burzovních indexů v náhodně vybraných historických časových obdobích.
2. S využitím statistických regresních metod získejte pro dané časové řady parametry regresních modelů s vhodně zvolenou strukturou.
3. Pomocí vybraných metod softcomputingu získejte parametry vhodných modelů popisujících stejné časové řady.
4. Analýzu opakujte pro různá časová období.
5. Jednotlivé modely ověřte na datech následujících po zvoleném časovém období. Porovnejte predikované hodnoty s historickou skutečností a zhodnoťte kvalitu jednotlivých modelů.

Rozsah diplomové práce:

Rozsah příloh:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

1. **MUSÍLEK, Petr.** Trhy cenných papírů. Praha: Ekopress, 2002, 459 s. ISBN 8086119556.
2. **DOSTÁL, Petr.** Soft computing v podnikatelství a veřejné správě. Brno: Akademické nakladatelství CERM, 2015, 2 sv. ISBN 9788072048960.
3. **CUESTA, Hector.** Analýza dat v praxi. Brno: Computer Press, 2015, 296 s. ISBN 9788025143612.
4. **FANTA, Jiří.** Psychologie, algoritmy a umělá inteligence na kapitálových trzích. Praha: Grada, 2001, 167 s. Finance. ISBN 8024700247.
5. **ZELINKA, Ivan, Zuzana KOMÍNKOVÁ OPLATKOVÁ a Roman ŠENKERÍK.** Aplikace umělé inteligence. Zlín: Univerzita Tomáše Bati ve Zlíně, 2010, 151 s. ISBN 9788073188986.
6. **NELLES, Oliver.** Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Berlin: Springer, c2001, xvii, 785 s. ISBN 3540673695.

Vedoucí diplomové práce: **doc. Ing. Marek Kubalčík, Ph.D.**

Ústav řízení procesů

Datum zadání diplomové práce: **3. prosince 2018**

Termín odevzdání diplomové práce: **15. května 2019**

Ve Zlíně dne 7. prosince 2018

doc. Mgr. Milan Adámek, Ph.D.  
*děkan*



prof. Mgr. Roman Jašek, Ph.D.  
*garant oboru*

**Prohlašuji, že**

- beru na vědomí, že odevzdáním diplomové práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou práci nebo poskytnout licenci k jejímu využití jen přípouští-li tak licenční smlouva uzavřená mezi mnou a Univerzitou Tomáše Bati ve Zlíně s tím, že vyrovnání případného přiměřeného příspěvku na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše) bude rovněž předmětem této licenční smlouvy;
- beru na vědomí, že pokud bylo k vypracování diplomové práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

**Prohlašuji,**

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor;
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně, dne 14.5.2019

Roman Smětač v.r.  
.....  
podpis diplomanta

## **ABSTRAKT**

V této práci je popsáno prostředí akciových trhů, včetně jejich klasických analýz. Popsány jsou také zdroje pro získání burzovních případně souvisejících ekonomických dat a případné problémy s nimi.

Jsou zde definovány časové řady a popsány metody pro jejich modelování. Zvoleny jsou také parametry pro modelování těchto řad.

Pro vyhodnocení kvality modelu a predikce jsou zvoleny vhodné funkce

Je porovnána predikce provedena jak pomocí statistických regresí, tak z oblasti softcomputingu (analytické programování).

Jeden směr tvorby predikce je jen na základě minulých hodnot, druhý směr predikuje hodnoty v závislosti na více vstupech.

**Klíčová slova:** predikce, burzovní data, autoregresní model, modelace časových řad, srovnání s analytickým programováním

## **ABSTRACT**

This work describes the stock market environment, including their classical analyzes. There are also described sources for obtaining stock exchange or related economic data and possible problems with them.

There are defined time series and methods for their modeling. Parameters for modeling these series are also selected.

Suitable functions are chosen to evaluate the model's quality and prediction.

The prediction is compared using both statistical regression and softcomputing (analytical programming).

One direction of prediction creation is based only on past values, the other direction predicts values depending on multiple inputs.

**Keywords:** prediction, stock exchange data, autoregressive model, time series modeling, comparison with analytical programming prediction

Chtěl bych na tomto místě poděkovat doc. Ing. Markovi Kubalčíkovi Ph.D. za přínosné podmínky a čas věnovaný při konzultacích.

Také děkuji své rodině za trpělivost a podporu po dobu celého studia.

# OBSAH

<b>ABSTRAKT</b> .....	<b>5</b>
<b>ABSTRACT</b> .....	<b>5</b>
<b>ÚVOD</b> .....	<b>9</b>
<b>I TEORETICKÁ ČÁST</b> .....	<b>10</b>
<b>1 POPIS ZKOUMANÉHO PROSTŘEDÍ</b> .....	<b>10</b>
1.1 ZÁKLADNÍ POJMY.....	10
1.2 ANALÝZY AKCIÍ.....	13
1.2.1 Fundamentální analýza.....	13
1.2.2 Technická analýza.....	14
1.2.3 Sezonnost akcií.....	17
<b>2 ČASOVÉ ŘADY</b> .....	<b>18</b>
2.1 STACIONARITA.....	18
2.2 DEKOMPOZICE ČASOVÝCH ŘAD.....	19
<b>3 STATISTIKA - REGRESNÍ MODELY</b> .....	<b>20</b>
3.1 METODA NEJMENŠÍCH ČTVERCŮ.....	20
3.2 AUTOREGRESNÍ (AR) MODEL.....	20
3.2.1 AR model řešený metodou nejmenších čtverců.....	21
3.2.2 Model AR (řád modelu a jeho úspěšnost).....	22
3.3 KLOUZAVÝ PRŮMĚR (MOVING AVERAGE).....	23
3.4 DALŠÍ REGRESNÍ MODELY POUŽITELNÉ PRO ČASOVÉ ŘADY.....	24
<b>4 SOFTCOMPUTING</b> .....	<b>24</b>
4.1 NEURONOVÉ SÍTĚ.....	25
4.2 GENETICKÉ PROGRAMOVÁNÍ.....	25
4.3 ANALYTICKÉ PROGRAMOVÁNÍ.....	25
<b>II PRAKTICKÁ ČÁST</b> .....	<b>26</b>
<b>5 PŘÍPRAVA DAT A POUŽITÉ PROSTŘEDKY</b> .....	<b>26</b>
5.1 VOLBA METOD.....	26
5.2 PARAMETRY A MOŽNÉ PROBLÉMY.....	26
5.3 ZDROJE DAT.....	27

5.4	PŘÍPRAVA DAT (PREPROCESSING).....	28
5.5	VYHODNOCENÍ KVALITY MODELU A PREDIKCE.....	28
<b>6</b>	<b>MODEL BEZ VSTUPŮ.....</b>	<b>29</b>
6.1	NEFILTROVANÁ DATA.....	30
6.1.1	Model.....	30
6.2	VYFILTROVANÁ DATA POMOCÍ KLOUZAVÝCH PRŮMĚRŮ.....	35
6.3	REGRESNÍ MODEL S VÍCE VSTUPY.....	36
6.3.1	Použitá data a jejich zdroje.....	36
6.3.2	Příprava dat.....	39
6.3.3	Tvorba modelu.....	39
	<b>ZÁVĚR.....</b>	<b>43</b>
	<b>SEZNAM POUŽITÉ LITERATURY.....</b>	<b>44</b>
	<b>SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....</b>	<b>46</b>
	<b>SEZNAM OBRÁZKŮ.....</b>	<b>47</b>
	<b>SEZNAM TABULEK.....</b>	<b>48</b>
	<b>SEZNAM PŘÍLOH.....</b>	<b>48</b>



## ÚVOD

Schopnost predikovat je jednou z významných schopností živočichů. Díky schopnosti odhadnout běh kořisti má dravec větší šanci ulovit potravu, při sportu hokejista odhadne směr a rychlost nahrávky spoluhráči, tak aby ji zpracoval, byl u ni dříve než soupeř, zkušený horal odhadne jaké bude počasí za pár hodin. Potřeba predikovat je v poslední době silnější díky větším dovednostem v oblastech matematiky a fyziky. Jedna z vět o predikci je: *„Predikce je činnost postavená na různých matematických algoritmech. Jejich cílem je co nejpřesněji odhadnout budoucí stav daného dynamického systému na základě současného stavu.“* [13]

Práce se zabývá možností predikce vývoje akciových titulů a její porovnání se skutečností pomocí klasických statistických regresních metod a také některým z prostředků softcomputingu, konkrétně v tomto případě analytickým programováním.

Použité metody z obou skupin se na zkoumaný systém dívají z vnějšku, tj. nesnaží se exaktně popsat vnitřní vztahy a nesnaží se najít všechny vstupy (včetně „poruch“) a důvody ovlivňující výstup (v tomto případě hodnotu akcií), ale zkoumají vnější chování systému.

Kromě toho byly predikovány hodnoty při použití regresní metody založené na závislosti více vstupech

Motivací pro vznik práce:

- byl dlouhodobý zájem o akciové trhy
- při studiu nové znalosti o přístupu k řešení problémů pomocí statistických a softcomputingových metod
- snaha si trochu více proniknout do oblastí mimo tradičních IT znalostí, tj. do oborů, které budou hrát čím dál větší roli (statistika, strojové učení, umělá inteligence)

# I. TEORETICKÁ ČÁST

## 1 POPIS ZKOUMANÉHO PROSTŘEDÍ

O prostředí burzy potažmo cenných papírů by se daly napsat knihy, ale toto není předmětem diplomové práce, tak budou uvedeny ve zkratce jen základní pojmy.

### 1.1 Základní pojmy

**Akcie** je v obchodním zákoníku České republiky popsána takto: „*Akcie je cenný papír nebo zaknihovaný cenný papír, s nímž jsou spojena práva akcionáře jako společníka podílet se podle tohoto zákona a stanov společnosti na jejím řízení, jejím zisku a na likvidačním zůstatku při jejím zrušení s likvidací.*“ [16] Jinými slovy, vlastnictvím akcií se stáváte poměrným spoluvlastníkem akciové společnosti a jako investor většinou očekáváte zhodnocení svých investic zvyšováním ceny akcie a případnou výplatou podílu na zisku (dividend). Kromě toho se můžete podílet jejím řízením (poměrný hlas na valných hromadách).

*Počátky trhů cenných papírů sahají k italským kupcům počátkem 15. století, obchody měly nestandardizovaný charakter a sekundární trh nebyl formálně organizován. [9]*

*Výraz **burza** pravděpodobně vznikl v Bruggách, kde se obchodníci s cennými papíry scházely v domě kupecké rodiny Van de Burse.[9]*

Nynější podobu světových trhů cenných papírů máme od posledních revolučních změn ke kterým došlo v 80. a 90. letech 20. století (globalizace finančních trhů, elektronizace a automatizace, zvýšení konkurence, atd.) díky kterým je v dnešní době umožněno obchodování široké mase lidí. V posledních letech se také zvyšuje podíl objednávek vygenerovaných algoritmy.

**Burza** je srdcem kapitálového trhu. Je to instituce, která organizuje tento trh pod zákonným dohledem.

Akcie se vydávají (emitují) při splnění předepsaných podmínek na **primárním trhu**, čímž emitent získá od investorů peníze na své podnikání. Toto je poměrně málo častá záležitost, ( u veřejně obchodovatelných akcií). Mnohem častější je jejich nákup či prodej na **sekundárním trhu** (např. tržní kapitalizace akcií na NYSE – New York Stock Exchange v dubnu 2017 byla  $23 \cdot 10^{12}$  USD a ten samý měsíc bylo obchodováno akcií za  $449 \cdot 10^9$  USD [nyxdata])

**Rozdělení akcií.** Akcie jsou rozděleny podle spousty kritérií do různých skupin. Nás budou zajímat pouze veřejně obchodovatelné akcie.

Kromě akcií se obchoduje podle stejných principů také s burzovními indexy, službami, komoditami jako je ropa, plyn, cukr, káva, bavlna, atd. včetně drahých kovů (zlato, stříbro, platina), měnovými páry, kryptoměnami (např. Bitcoin, Ethereum, Ripple)

**Účastníci burzy.** Účastníky burzy jsou subjekty rozdělené do základních skupin:

- poskytovatelé kapitálu (např. soukromí investoři, fondy, penzijní fondy,..)
- subjekty poptávající kapitál ( firmy, veřejný sektor)
- zprostředkovatelé (zprostředkovatelé, obchodníci, poradci) tzv. členové burzy
- zákonné a dozorové instituce (např. u nás Ministerstvo financí, ČNB)

**Deriváty, futures.** Kromě klasických akcií (popř. indexů, komodit atd.) se také obchoduje s deriváty těchto aktiv (CFD – Contract for difference, waranty, certifikáty) kdy není třeba platit za celou hodnotu akcie.

Tyto deriváty, futures mají jako podkladové aktivum akcie (popř. indexy, komodity,..) a jejich cena se odvíjí od ceny tohoto aktiva.

U derivátů se výrazně uplatňuje **finanční pákový efekt** jak při kladném, tak při záporném zhodnocení. Například při nákupu 100 CFD na akcii s pákou 1:10, kdy hodnota jedné akcie je 1000Kč je skutečná investice jen 10 000 Kč (100ks po 100Kč) místo 100 000 Kč. Je také třeba počítat se započtenou cenou do pořízení a rozdílu prodejní a nákupní ceny. Pokud se při prodeji cena akcie změnila o 5%, tak se cena CFD změnila o 50%, tj. zisk nebo ztráta 50% investované částky (opět bez započítání případné režie).

**Dividendy** jsou podíl na zisku vyplácený některými společnostmi. Akcie, které pravidelně a v poměrně příznivém poměru k aktuální tržní ceně vyplácejí dividendy bývají vyhledávány konzervativnějšími investory. Několik z titulů v České republice má parametry těchto dividendových akcií a má poměrně slušný dividendový výnos v poměru k aktuální ceně. Např. v roce 2018 ČEZ: 5,4%, Moneta: 8%, O2: 6,3%, Philip Morris ČR: 5,3% (vše počítáno v čisté ceně po odečtení 15% srážkové daně a vztažené k závěrečným cenám akcií dne 3.4.2018).[5]

**Ex-divident day.** Nejdůležitější datum, které sledují dividendoví investoři. První den, kdy lze prodat akcie bez ztráty nároku na výplatu dividendy nebo také první den, kdy kupující již nemá nárok na nejbližší dividendu. Pro získání dividendy musíte tedy akcie koupit PŘED ex-dividend date. Pokud koupíte akcie v ex-dividend date, již nebudete mít nárok na nejbližší dividendu.[7]

**Prodej na krátko (short).** Dá se také spekulovat na pokles ceny akcií (derivátů,atd.). Principem je zapůjčení si akcií od držitele akcií (za poplatek) a jejich okamžitý prodej. V okamžiku, kdy zapůjčené akcie vrátíte si je za aktuální cenu nakoupíte na trhu a pokud je aktuální cena nižší, vyděláte na rozdílu.

**Kurzovní riziko.** Při investicích do akcií v jiných měnách je nutné brát v potaz kurzovní riziko. Proti tomuto se dá zajistit např. nákupem derivátu na opačný kurzovní vývoj.

**Burzovní indexy.** Index cen vybraných akciových titulů. Dává přehled o vývoji a stavu trhů, případně odvětví. Nejčastěji je index sestavován z nejkvalitnějších akcií a podíl vlivu cen na tento index je často vážen (podle kapitalizace, cen). Investicí do indexů je obvykle méně rizikové nežli investice do samostatných akcií. Výkonnost investičních fondů je často porovnávána s výkonností indexů. Znamé indexy jsou např. Dow Jones Industrial Average, Nasdaq, Standard and Poor's 500, DAX, atd.

**Investování.** Ve světě (západním), je dlouhá tradice investování do akciových titulů a jsou částí rodinných investic. Časový horizont investic je spíše delší (investiční fondy doporučují cca 5 let, jednotlivé investice v řádech měsíců). U kratších horizontů (dny, hodiny tiky) jde spíše o spekulaci či hru. Riziko investování do akcií je vysoké, vysoká je ale také možnost potenciálních výdělků. Rozpětí rizika je velmi značné. Od nejméně rizikových např. některých typů certifikátů, přes indexy či Blue chips (akcie největších, nejčastěji obchodovaných a nejlépe hodnocených společností), až po deriváty, či tzv. centové akcie.

## 1.2 Analýzy akcií

K analýzám akcií se používají převážně všeobecně uznávané metody oceňování a to **fundamentální analýza** a **technická analýza**. Obě metody mají své příznivce i odpůrce a navzájem si vyčítají své nedostatky.

Často se obě metody kombinují a podle mého názoru má fundamentální analýza větší vliv u delších horizontů rozhodování (měsíce a déle) a technická spíše na kratší.

Kromě těchto analýz investoři musí kalkulovat také s **psychologickými vlivy** (které se také analyzují) a množstvím různých předvídatelných i nepředvídatelných událostí majících výrazný vliv na pohyb cen (války, politická rozhodnutí, ekonomické ukazatel, nové objevy, vliv počasí, stav zásob, atd.).

Událostí mívají obvykle téměř okamžitý a často přehnaný vliv na kurz i když předvídaný ekonomický dopad nastane až ve mnohem vzdálenější době.

### 1.2.1 Fundamentální analýza

Fundamentální analýza předpokládá, že každá akcie či aktivum má svou vnitřní hodnotu, která vychází z historických dat a hospodaření společnosti a aktuální kurz akcie se pohybuje kolem této vnitřní hodnoty. Analýza se nejdříve snaží za použití finančních analýz zjistit vnitřní hodnotu a prognózovat budoucí vývoj.

Často zmiňovaný údaj je P/E (cena akcie/zisk). Znamená počet let, za který společnost vyprodukuje při současné výkonnosti zisk rovnající se kapitalizaci.

Tato analýza není předmětem diplomové práce.

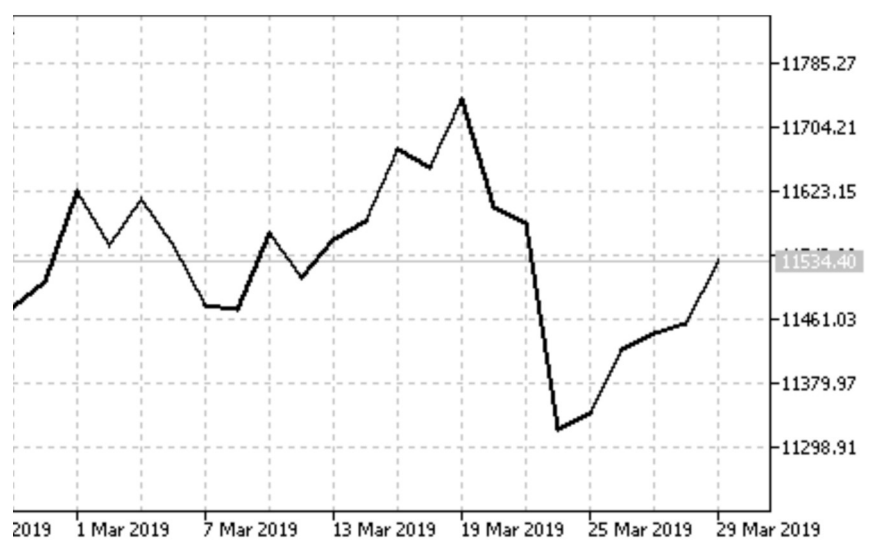
### 1.2.2 Technická analýza

Technická analýza vychází z přesvědčení, že všechny vlivy na vývoj analyzované akcie jsou zahrnuty v grafu. Nic jiného již není potřeba a veškerá rozhodnutí se dají vyčíst z něj. Snahou technické analýzy je odhalit kdy je aktivum levné nebo naopak předražené, odhalení trendů, hlavně jejich změnu.

#### Grafické znázornění

Základem technické analýzy je grafické znázornění kurzu aktiva za zvolená období a zvoleným vzorkováním. Nejčastěji spojuje se znázorněním indikátorů, ale někteří investoři se řídí pouze svým instinktem na základě vývoje grafu.

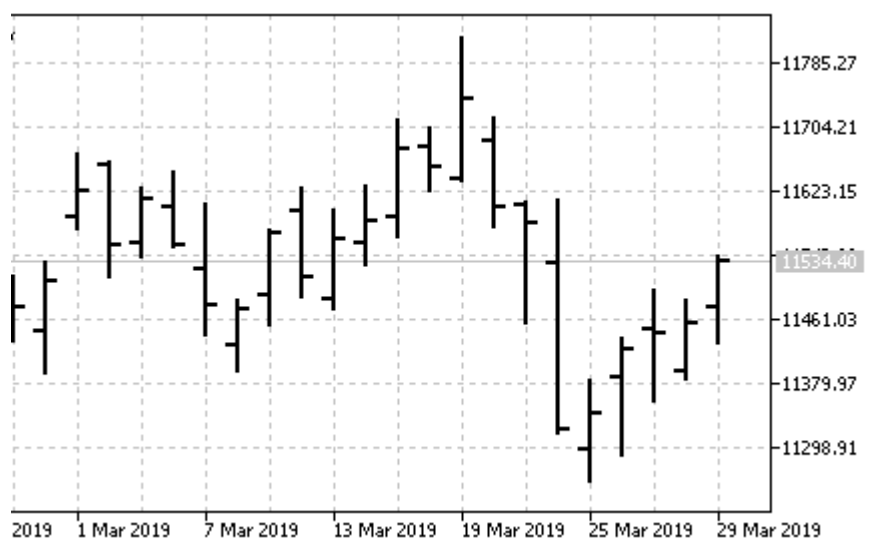
Jednodušší grafy zobrazují pouze **koncové kurzy**



Obr. 1: Zobrazení jen koncových kurzů

Často používané typy grafů, které znázorní více údajů (vstupní, výstupní, maximální a minimální) jsou:

**sloupcové** (horizontální čárky jsou vstupní a výstupné hodnoty a vertikální sloupeček rozsah hodnot)



Obr. 2: Sloupcový graf

**svíčkové** (svíčka ohraničuje vstupní a výstupní hodnotu, horní a dolní knot dosažená maximální či minimální dosažená hodnota v rámci vzorku. Často bývají rozlišeny svíčky pro stoupající a klesající hodnoty plnou či prázdnou plochou, popř. barvami)



Obr. 3: Svíčkový graf

Kromě vlastních hodnot zajímají technické analytiku také zobchodované **objemy**.

### Rozdělení

Klasické nástroje technické analýzy můžeme dělit do dvou základních skupin:

- formace v grafu
- technické indikátory

**Formace v grafu.** V grafu se vyhledávají (většinou po vyhlazení vývoje pomocí klouzavého průměru) grafické tvary, které dávají předpoklad dalšího vývoje. Příklady formací:

- dvojitý vrchol a dvojitě dno
- hlava rameno
- kulatý vrchol, kulaté dno
- podpora a odpor
- trendové čáry a kanály
- trojúhelníky
- vlajková formace
- doji star, morning star, a podobná formace svíčkových grafů

Zvláštní formou grafických metod je **Elliotova teorie vln**. Základem je teorie, že pohyb akciových kurzů hlavním směrem je přerušován slabšími protisměrnými pohyby. Impulzivní vlny sledují všeobecný trend a skládá se z 5 vln. Korekturní vlny se pohybují proti trendu a skládají se ze 3 vln. [9]

**Technické indikátory, oscilátory.** Je to několik desítek a jsou konstruovány na základě minulých cen podkladového aktiva a objemů uskutečněných obchodů. Jsou to vzorce, jejichž výpočtem dostáváme informace o tom, že je některé aktivum překoupené nebo přeprodané a také o tom, že cena již změnila směr, což může být začátek nového silného trendu.

Technické indikátory lze zobrazit graficky a většinou mají podobu křivky oscilující



kolem neutrální linie. [17]

Mezi pravděpodobně nejznámější a nejpoužívanější patří:

- kombinace klouzavých průměrů
- Momentum
- MACD (Moving Average Covergence Divergence)
- RSI (Relative Streght Index)
- Bollinger bands
- Stochastic oscilátory
- W%R (Williams Percentage Range)[4]

### 1.2.3 Sezonnost akcií

O sezónnosti akcií existuje spousta úsloví. K nejznámějším asi patří „ Sell on May and go away“ (prodej v květnu – do konce října nic nepodnikej) či tzv. „Lednový efekt“ - kdy podle tohoto tvrzení intenzivněji se zvednou ceny během ledna. Existují různé studie o jejich platnosti či neplatnosti, ale je to podobné jako v našich podmínkách úsloví o počasí (ledoví muži, Medardovo kápě apod.) - zná je téměř každý.

## 2 ČASOVÉ ŘADY

Pod pojmem časové řady rozumíme chronologicky uspořádaná data (pozorování). Intervaly pozorování jsou zpravidla rovnoměrné (ekvidistantní). Podle okolností to mohou být zlomky sekund, minuty, hodiny, čtvrtletí, roky.

*Cílem analýzy časové řady je určení modelu (mechanismu), podle něhož jsou generována sledovaná data. Znalost tohoto modelu umožňuje předpovídat budoucí vývoj systému a do jisté míry i řídit a optimalizovat chování systému vhodnou volbou vstupních parametrů a počátečních podmínek.[6]*

K analýze časových řad jsou nejčastěji používány následující metody:

- dekompozice časových řad
- Boxova-Jenkinsonova metodologie
- lineární kauzální modely
- spektrální analýza časových řad (Fourierovy analýzy)

### 2.1 Stacionarita

Stacionarita je obvyklým předpokladem většiny technik analýzy časových řad.

Jedná se o náhodný proces, jehož rozdělení pravděpodobnosti se v čase nemění. V důsledku toho se nemění ani parametry jeho pravděpodobnostní funkce (např střední hodnota, rozptyl). Autokorelační funkce stacionárního procesu závisí pouze na rozdílu svých argumentů.

Předpokladem stacionarity se rozumí ty časové řady či signály, které jsou bez trendu, mají s měnícím se časem stejný rozptyl a stejnou podobu autokorelační funkce. [Schwartz]

V případě nestacionárních časových řad lze provést:

- diferencování  $dx_i = x_i - x_{i-1}$

- odstranění trendu odečtením proloženého polynomu
- stabilizace rozptylu logaritmizací čtverce řady

## 2.2 Dekompozice časových řad

Časové řady mohou být rozloženy na několik složek:

- trend
- sezonní složku
- cyklickou složku
- náhodnou (reziduální složku)

**Trend** odráží dlouhodobé změny v průměrném chování. Je způsoben jevy, které působí jedním směrem.

**Sezonní složka** zahrnuje jevy, které se dějí se během jedné sezóny a opakující se každou další sezónu.

**Cyklická složka** je střídání fází růstu a poklesu kolem trendu.

**Náhodná složka** zahrnuje jevy které nemají systematický charakter. Předpokládá se, že náhodná složka má charakter **bílého šumu** ( náhodné veličiny s nulovou střední hodnotou konstantním rozptylem)

Dekompozice těchto složek ( $Tr$  – trend,  $Sz$  – sezonní složka,  $C$  – cyklická složka a  $\varepsilon$  – náhodná složka) je dvojího typu:

- aditivní

$$Y_t = Tr_t + Sz_t + C_t + \varepsilon_t$$

- multiplikativní

$$Y_t = Tr_t \cdot Sz_t \cdot C_t \cdot \varepsilon_t$$

Za využití metod regresní analýzy se pracuje pouze se systematickými složkami (trend, sezonnost, cyklická složka)

### 3 STATISTIKA - REGRESNÍ MODELY

Statistika jako věda má kromě **deskriptivní (popisné) funkce** mnoha jevů pro kterou má definovanu řadu pojmů (průměr - střední hodnota, medián, modus, rozptyl, směrodatná odchylka (odmocnina rozptylu), atd.) zabývá také s určitou pravděpodobností **předvídáním jevů** na základě jejich předchozích stavů, či vlivů jiných jevů a veličin.

Postupem času je měřeno, ukládáno stále větší a větší množství dat. Tato big data se velmi často dále zpracovávají a vyhodnocují statisticky a význam statistiky roste.

Jak řekl v roce 2009 Hal Varian, vedoucí ekonom společnosti Google: „*Neustále říkám, že nejzajímavější prací příští dekády bude statistika. Lidé si myslí, že žertuji, ale kdo by si býval myslel, že v devadesátých letech bude nejzajímavější prací programování?*“ [2]

#### 3.1 Metoda nejmenších čtverců

Obecně metoda nejmenších čtverců matematicko-statistická je metoda, sloužící k minimalizaci chyb k pevně danému kritériu, tj. nalezení řešení takového, aby součet druhých mocnin chyb byl co nejmenší. [15] .

#### 3.2 Autoregresní (AR) model

Autoregresní model časových řad vychází z toho, že každá hodnota časové řady je v relaci s předchozími hodnotami

AR model řádu N je definován:

$$y(k) = -c_1 y(k-1) - c_2 y(k-2) + \dots - c_N y(k-N) - e(k) \quad (1)$$

kde  $y(k)$  ..... je nová hodnota vypočtená na základě předchozích hodnot  
 $c_1, c_2, \dots, c_N$  ... koeficienty autoregresního procesu  
 $e(k)$  ..... chyba (současná hodnota), kterou je aproximován bílý  
šum

[3]

### 3.2.1 AR model řešený metodou nejmenších čtverců

$$\begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} -y(n) & -y(n-1) & \dots & -y(1) \\ -y(n+1) & -y(n) & \dots & -y(2) \\ \vdots & \vdots & \ddots & \vdots \\ -y(N-1) & -y(N-2) & \dots & -y(N-n) \end{bmatrix} \begin{bmatrix} c(1) \\ c(2) \\ \vdots \\ c(n) \end{bmatrix} + \begin{bmatrix} \hat{e}(1) \\ \hat{e}(2) \\ \vdots \\ \hat{e}(n) \end{bmatrix} \quad (2)$$

rovnici lze přepsat na tvar

$$y = F \hat{\Theta} + \hat{e} \quad (3)$$

,kde

$$y = \begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(N) \end{bmatrix} \quad (4)$$

,

$$\hat{e} = \begin{bmatrix} \hat{e}(1) \\ \hat{e}(2) \\ \vdots \\ \hat{e}(N) \end{bmatrix} \quad (5)$$

a

$$F = \begin{bmatrix} \phi^T(1) \\ \phi^T(2) \\ \vdots \\ \phi^T(N) \end{bmatrix} \quad (6)$$

a z ní vyjádřit chybový vektor

$$\hat{\mathbf{e}} = \mathbf{y} - F \hat{\Theta} \quad (7)$$

Po volbě kritéria minimalizace kvadrátu odchylky

$$J = \hat{\mathbf{e}}^T \hat{\mathbf{e}} \rightarrow \min \quad (8)$$

a použití vztahu

$$\frac{\partial(\mathbf{x}^T)\mathbf{z}}{\partial \mathbf{a}} = \frac{\partial \mathbf{x}^t}{\partial \mathbf{a}} \mathbf{z} + \frac{\partial \mathbf{z}^T}{\partial \mathbf{a}} \mathbf{x} \quad (9)$$

se rovnice derivuje

$$\begin{aligned} \frac{\partial J}{\partial \hat{\Theta}} &= \frac{\partial \hat{\mathbf{e}}^T \hat{\mathbf{e}}}{\partial \hat{\Theta}} = \frac{\partial (\mathbf{y} - F \hat{\Theta})^T}{\partial \hat{\Theta}} (\mathbf{y} - F \hat{\Theta}) + \frac{\partial (\mathbf{y} - F \hat{\Theta})^T}{\partial \hat{\Theta}} (\mathbf{y} - F \hat{\Theta}) \\ \frac{\partial J}{\partial \hat{\Theta}} &= -F^T (\mathbf{y} - F \hat{\Theta}) - F^T (\mathbf{y} - F \hat{\Theta}) = -2F^T (\mathbf{y} - F \hat{\Theta}) = 0 \end{aligned} \quad (10)$$

Výsledný vektor odhadu parametrů je

$$\hat{\Theta} = (F^T F)^{-1} F^T \mathbf{y} \quad (11)$$

### 3.2.2 Model AR (řád modelu a jeho úspěšnost)

Řád modelu určuje řád polynomu, v podstatě počet hodnot jdoucích do výpočtu. V praxi se nejčastěji používá řád 2.

Úspěšnost modelu se může vyjádřit indexem spolehlivosti daným vzorcem (12). Jeho hodnota by měla být mezi 0 až 1. Vyšší hodnota je lepší

### 3.3 Klouzavý průměr (Moving Average)

Klouzavý průměr je metoda k odhalování trendů v časových řadách a díky tomu jednou z nejpoužívanějších metod v technické analýze. K jeho oblíbenosti v technické analýze přispívá také jeho snadný výpočet a možnost kombinovat několik klouzavých průměrů navzájem. Jeho nevýhodou je jisté zpoždění za aktuálním vývojem.

U klouzavého průměru se na základě lineární kombinace konstruuji hodnoty v minulosti k čemuž se používá různých koeficientů. Dochází k vyhlazení prudkých změn hodnot, či chceme-li jazykem teorie signálů k filtraci vysokých frekvencí popř odstranění šumu.

Pravděpodobně nejpoužívanější klouzavé průměry jsou:

- **Jednoduchý klouzavý průměr (Simple Moving Average)**

Jde o aritmetický průměr posledních konstantních  $N$  pozorování. Délka  $N$  pozorování se volí na základě charakteru řady či empirických výzkumů a také určují jeho stupeň a označení (např. posledních 20 hodnot je SMA20, posledních 100 hodnot SMA100)

Jeho výpočet je:

$$\bar{y}_k = \frac{1}{N} \sum_{i=0}^{N-1} y_{k-i}, \text{ kde } N \text{ je délka průměru}$$

- **Exponenciální klouzavý průměr (Exponential Moving Average)**

Tento klouzavý průměr na rozdíl od jednoduchého dává vyšší váhu posledním údajům.

Jeden z možných výpočtu je:

pro  $k=1$

$$EMA_k = y_1$$

pro  $k > 1$

$$EMA_k = N \cdot y_k + (1 - N) \cdot EMA_{k-1}$$

- **Vážený klouzavý průměr (Weighted Moving Average)**

Alespoň jedné hodnotě je změněna váha jdoucí do výpočtu. Obvykle se jedná o vyšší váhy posledních hodnot. Speciálním případem je triangulární klouzavý průměr (Triangular Moving Average), kdy největší váhu má prostřední hodnota.

Rozdíl použití klouzavého průměru se liší u technické analýzy a u popisné statistiky. U **technické analýzy** se používá průměr z předchozích hodnot včetně poslední známé (hodnoty budoucí neznáme)

V **popisné statistice** se používají k vyrovnání řad hodnoty v okolí vyrovnávaného bodu tj. do průměru je zahrnut stejný počet hodnot minulých i budoucích včetně aktuální. Téměř výhradně je použit lichý řád.

[14]

[8]

### 3.4 Další regresní modely použitelné pro časové řady

Jako dalších modelů pro modelování časových řad se používají:

- MA (moving average)
- ARMA (kombinace AR a MA modelu)
- ARIMA (navíc přidaná integrační složka)

Další modely jsou víceméně teoretické.

## 4 SOFTCOMPUTING

Tato vědní disciplína se zabývá symbiózou různých výpočetních postupů jejichž společným jmenovatelem je odklon od klasického modelování založeného na booleovské logice, analytických modelech, ostré klasifikaci a deterministickém prohledávání. [11] k řešení obtížných problémů

Jako vhodné techniky pro řešení problémů časových řad jsou:

- neuronové sítě
- genetické programování
- analytické programování



## 4.1 Neuronové sítě

Poměrně dobré výsledky dávají neuronové sítě u vícevstupové regrese. Neuronové sítě jsou schopny úspěšně aproximovat i komplikované funkce.

Další možnost je odhalovat formace v grafech [13]

## 4.2 Genetické programování

Genetické programování je rozšířením genetických algoritmů. Tato modifikace je, že není nová populace šlechtěna klasickým numerickým přístupem, ale na symbolické úrovni. Tzn. Populace neobsahuje čísla, ale symbolické objekty (matematické funkce, uživatelské podprogramy) [12]

## 4.3 Analytické Programování

Analytické programování je nový přístup k syntéze symbolické struktury, který pro výpočet používá evoluční algoritmus. Od jeho zavedení se v mnoha problémech osvědčil jako vhodný pro syntézu symbolické struktury jako genetické programování. [10]

Jedná se o alternativní přístup vzhledem ke genetickému programování a gramatické evoluce. Analytické programování není samostatný evoluční algoritmus program, ale spíše transformace či zobrazení základních symbolických objektů do množiny programů, které lze z těchto symbolických objektů zkonstruovat.

Analytické programování pracuje stejně jako genetická evoluce nebo genetické programování s množinou funkcí (sin, tan, cos, and, or), operátorů (+, -, \*, /, dt), terminálů (2, 73,  $\pi$ , t,..) [12]

## II. PRAKTICKÁ ČÁST

### 5 PŘÍPRAVA DAT A POUŽITÉ PROSTŘEDKY

#### 5.1 Volba metod

Na vývoj cen akciových titulů má vliv obrovské množství vstupů. Tyto vstupy se v poslední úrovni sejdou v nabídce a poptávce. Nabídku a poptávku vytváří množství investorů kteří se řídí zase svými individuálními vstupy (výsledky analýz technických analýz, předchozím vývojem, záměrem vlastnictví určité části společnosti, psychologie a spoustou dalších podnětů). Protože těchto vstupů je obrovské množství a nejsou v měřitelné (možno považovat za **bílý šum**), tak se v **první fázi** se dívám na vývoj cen jako na **systém bez vstupů**, tj. čistou časovou řadu.

#### 5.2 Parametry a možné problémy

Možné problémy:

- **Přeučení** (overfitting) a **nedoučení** (underfitting). Protože zdrojová data jsou značně stochastická, přílišná snaha navrhovaného řešení o přizpůsobení se trénigovým datům na úkor obecnosti vede k přeučení. U AR modelů může vzniknout vysokým stupněm polynomu na malém množství dat, u neuronových sítí vysokým množství skrytých neuronů. Naopak může dojít k **nedoučení**, pokud bude model špatně nastaven (nízký stupeň polynomu, nízký počet skrytých neuronů) a predikce bude velmi slabě kopírovat vývoj, případně nebude fungovat vůbec.
- Data budou tak stochastická, že modely nebudou mít lepší výsledky, nežli pravděpodobnost ze statistického popisu dat.
- Bude zvoleno nevhodné vzorkování ( u akcií obvykle od ticků, přes minuty, násobky minut, hodiny, dny, týdny až měsíce)
- Ve výpočtech nebudou zahrnuty velké výkyvy mezi jednotlivými vzorky (občas se stane, že v rámci dne se ceny v reakci na nějakou významnou událost změní o více než předchozí změny za několik týdnů a stačí se vrátit zpět téměř na hodnoty před událostí).
- Obecně bývají také problémy s kalendářem. Různé délky dnů v měsíci/v roce, počtu pracovních dnů, pohyblivé svátky.

### 5.3 Zdroje dat

Zdarma veřejně dostupných zdrojů historických dat na denní bázi (vzorkování jeden den) je poměrně hodně. Často používaný zdroj je stránka společnosti **Yahoo** (<https://finance.yahoo.com>), kde po zadání zkratky požadovaného produktu a zvolený časový rozsah, je možné získat ve formátu csv.

Trochu problematictější je získat data s kratším vzorkováním. Obvykle jsou poskytovány zprostředkovateli svým zákazníkům (já mám k dispozici v programu MetaTrade historická data s možností exportu do csv se vzorkováním 1 minuta, 5 minut, 15 minut, 30 minut až měsíc). Pokud zůstanu u veřejně dostupných zdrojů tak velmi dobrý zdroj je ruský web **Finam** (<https://www.finam.ru/analysis/profile041CA00007/default.asp>), kde lze stáhnout

opět ve formátu csv data se vzorkováním už od jednotlivých tiků (tick). Americké akcie jsou při volbě skupiny *Акции США(BATS)*, futures amerických indexů *Фьючерсы США*.

Pro další analýzu byla vybrány údaje indexu Dow Jones Industrial Average a akcie firmy IBM, oboje s denním vzorkováním (od 1.1.2000 až po 29.3.2018) a 15 minutovým vzorkováním.

#### 5.4 Příprava dat (preprocessing)

Protože se jedná o kvalitní zdroj dat a v klasickém formátu (csv soubor ASCII kódování), nebylo nutno žádných úprav. Jen při zkoušení různých akcií a indexů jsem narazil na drobný problém a to hodnoty **null** u některých neobchodních dnů. Toto jsem řešil **odstraněním** takovýchto záznamů.

Data jsou uspořádána v řádcích chronologicky po jednotlivých vzorcích (vzorkování: dny, 15-minuty) a ve sloupcích, kde kromě hodnoty vzorku jsou hodnoty otevírací (není stejná s předchozí zavírací), maximální, minimální, zavírací, upravená zavírací a objem. Pro další zpracování jsem ponechal anglické názvy (Date, Time, High, Low, Close Adj Close, Volume).

Důvodem sloupce *upravená zavírací hodnota* (Adj Close) je používání těchto dat pro analýzy. Při významných událostech kdy např. dojde k rozdělení akcií (split) nebo výplatě dividendy, se cena akcie se výrazně změní. Tím, že je hodnota upravena (o změnu způsobenou událostí) nezkrusí to případné analýzy. U indexů je tato změna u jednotlivých akciových titulů zahrnutých v indexu zanedbatelná včetně vlivu na index a obě ceny se rovnají.

Data jsem používal kontinuálně, tj. řada pokračovala ihned dalším časem/dnem po po případném neobchodním období.

#### 5.5 Vyhodnocení kvality modelu a predikce

Jako míru kvality použitého modelu jsem použil **index determinace  $R^2$**  nazývaný také **index spolehlivosti**

$$R^2 = 1 - \frac{\sum_{k=1}^n (m_k - y_k)^2}{\sum_{k=1}^n (y_k - \bar{y})^2} \quad (12)$$

kde  $y_t$  je aktuální (známá) hodnota,

$m_k$  je modelovaná hodnota

$\bar{y}$  průměr známých hodnot

Hodnoty  $R^2$  blízké 1 znamenají kvalitní model, blízké 0 málo kvalitní (možno nahradit střední hodnotou).

Jako míru kvality použitého modelu i předpovědi jsem použil **průměrnou absolutní procentuální odchylku MAPE** (Mean Absolute Percentage Error), která dle mého názoru nejlépe a nejnázorněji vystihuje úspěšnost (jen pozor, aby se nepoužívala u dat, která mají případnou nulovou hodnotu). Její výhoda je nezávislost na měřítku řady.

$$MAPE = \frac{100}{n} \sum_{k=1}^n \frac{|y_k - \hat{y}_k|}{y_k} \quad (13)$$

kde  $y_k$  je aktuální (známá) hodnota,

$\hat{y}_k$  je predikovaná popř. modelovaná hodnota .

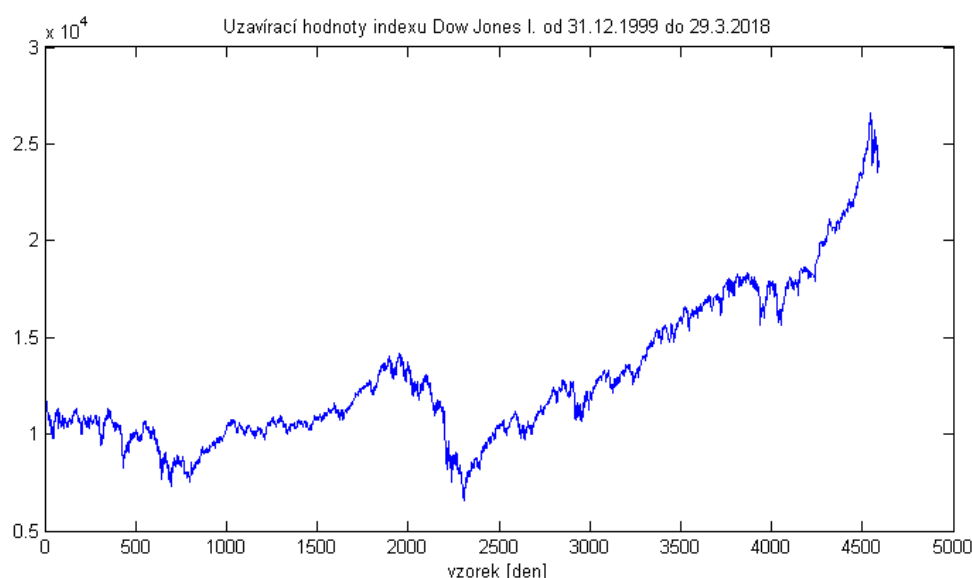
Kromě této vybrané míry kvality se běžně používají další, např. součet čtvercových chyb (SSE), střední čtvercová chyba (MSE), odmocněná střední čtvercová chyba (RMSE), střední absolutní chyba (MAE), relativní absolutní chyba (RAE), odmocněná relativní čtvercová chyba (RRSE)

Protože změny v krátkých intervalech a hlavně u indexů jsou často poměrně malé, byla testována úspěšnost predikce také porovnáním reálné hodnoty s vůči tomu, že není provedena žádná změna, případně vůči průměrné změně.

## 6 MODEL BEZ VSTUPŮ

## 6.1 Nefiltrovaná data

Jako první pro analýzu byl vybrán Dow Jonesův index se vzorkováním jeden den. Zvolil jsem období od 31.12.1999 po 29.3.2018. Protože většinou budu pracovat pouze s pořadovým číslem v rámci řady, bylo ponecháno na ose x pořadové číslo.



Obr. 4: Dow Jones Industrial Average – celá datová sada

### 6.1.1 Model

Podstatné pro AR model je volba řádu a množství dat zahrnutých do výpočtu. Jako výchozí parametry se zdálo vytvořit kvalitní model a poté změnou parametrů v hledat v testovací fázi úspěšného kandidáta. Úspěšnost predikce bude vůči skutečným pro model neznámých dat a ne míru úspěšnosti modelu na datech ze kterých byl vytvořen model.

V prvním kroku byla snaha vytvořit model, který by dostatečně dobře proložil zkoumaná data. Ideálem bylo najít kompromis mezi tím, aby proložení modelu bylo dostatečně kvalitní ( $R^2$  parametr blízký 1) a ne příliš vysokým stupněm AR. Alternativní varianta se také jevila postupně zvyšovat na stejných datech stupeň modelu od nejnižšího a porovnávat vypočtený index spolehlivost dle rovnice (12). V případě, že by jeho hodnota stagnovala, popř. se zvyšovala s rostoucím stupněm nepatrně, mohl

to být vhodným výchozím nastavením pro další testování predikcí. Kvalita modelu sice nic nevyovídá o úspěšnosti predikce, ale stupeň AR se může stát výchozím pro další testování predikcí.

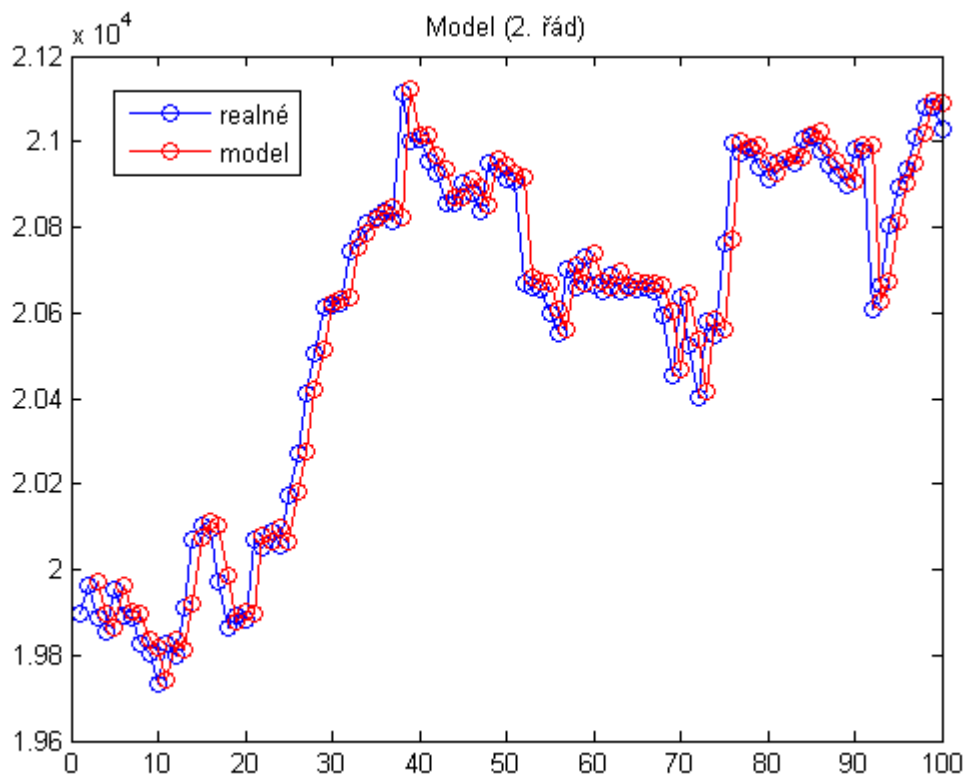
Dle doporučení se začalo v praxi často používaným stupněm AR modelu 2 a 100 hodnot zahrnutých do výpočtu.

Jako vzorový příklad je zvolen Dow Jonesův index

Po importu do MATLABu (pomocí skriptu Import\_DJ\_den) a volbě dat pro model (100 hodnot tj. 4281 až 4380 ) byly vypočítány pomocí metody nejmenších čtverců koeficienty (pomocí funkce mujAR )

$$c_1 = -0.9821 \quad c_2 = -0.0184 \quad (14)$$

Tomu odpovídá graficky znázorněný model

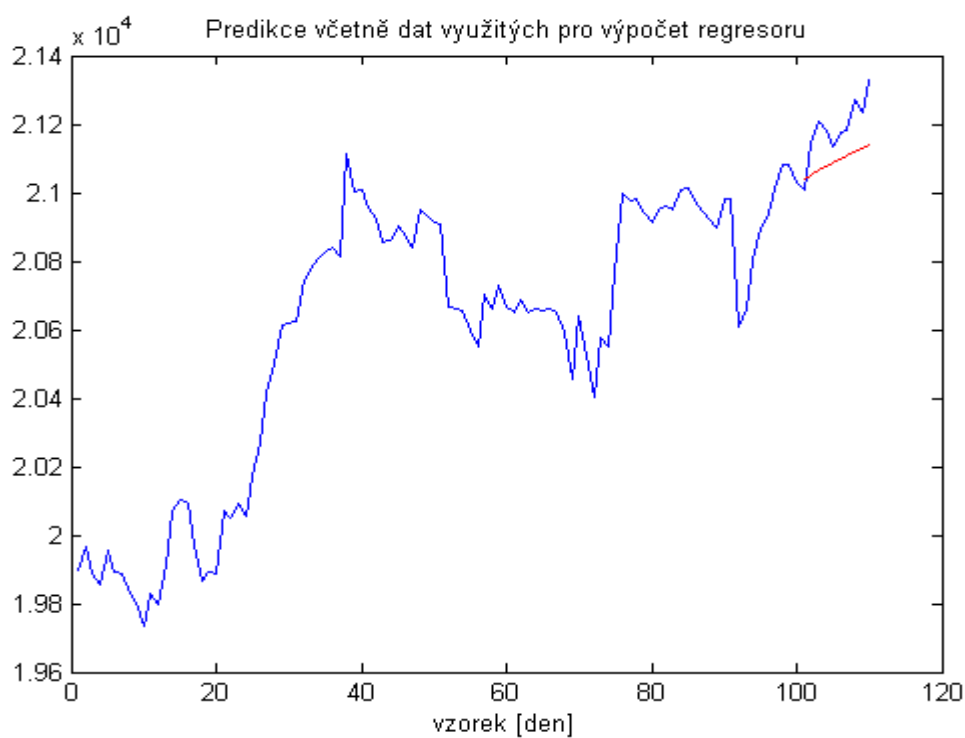


Obr. 5: Vzorový AR model DJI denní vzorkování

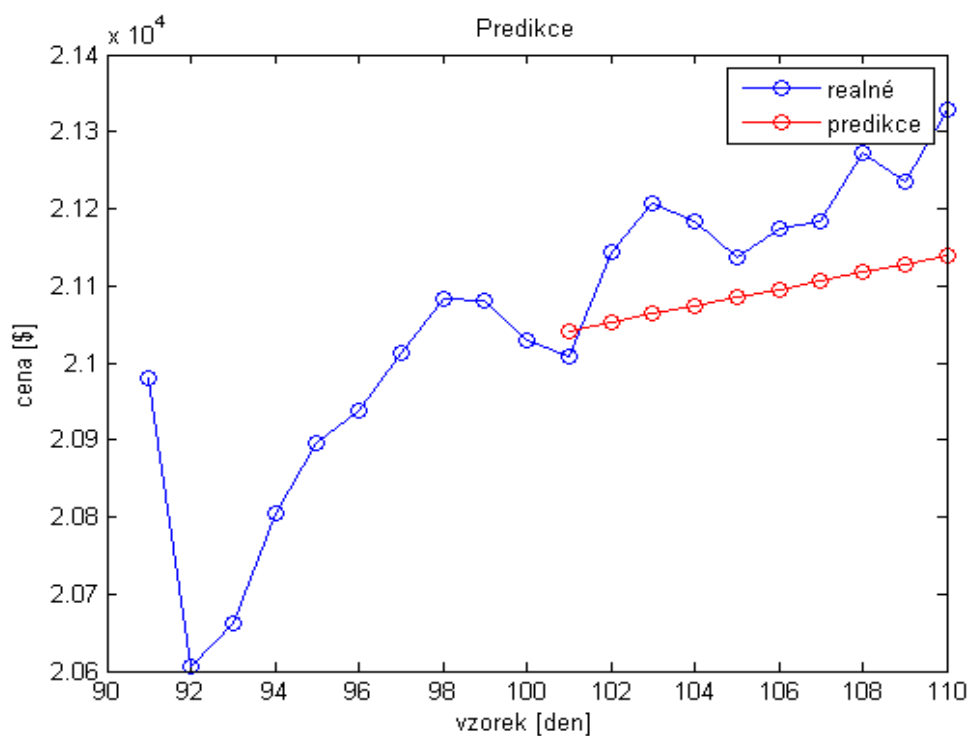
jeho index spolehlivosti dle vzorce (12) je

$$R^2 = 0,9448 \quad (15)$$

První predikovaná hodnota 4381 (odpovídající dnu 1.6.2017)



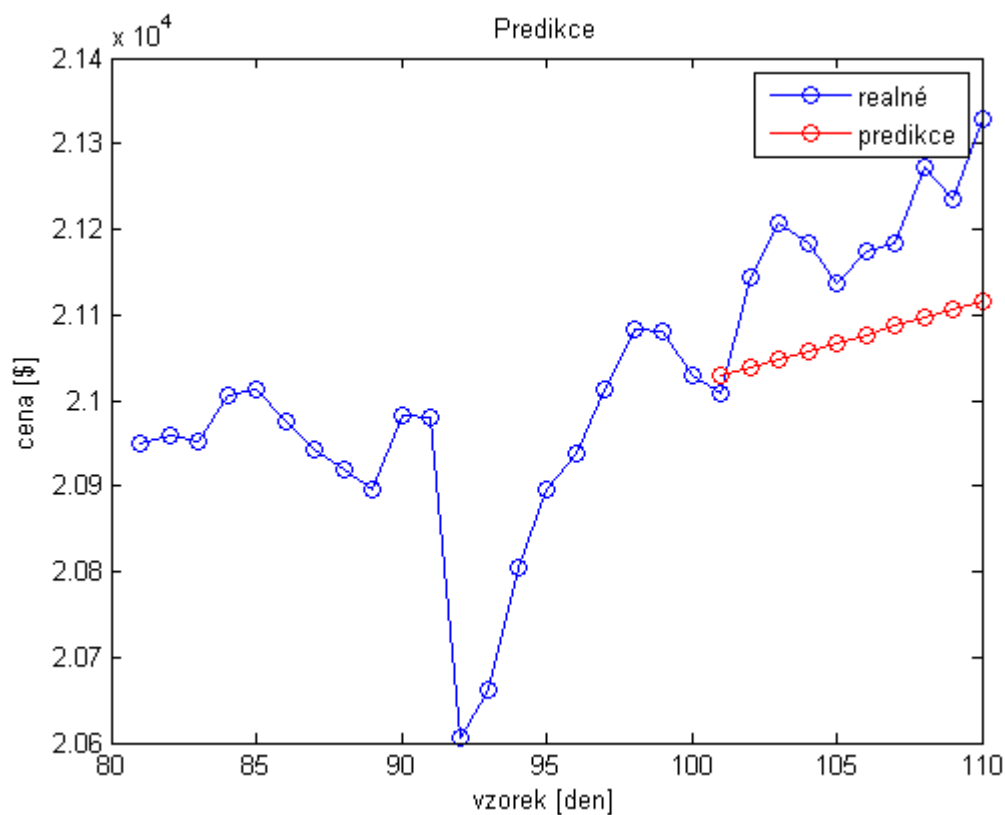
Obr. 6: Vzorový AR model DJI predikce komplet



Obr. 7: Vzorový AR model DJI detail predikce



Obdobně dopadlo analytické programování:



Obr. 8: Vzorový model DJI detail genetické programování

Tab. 1. Výsledky predikce vzorový DJI

<b>Reálné hodnoty</b>	<b>Genetické prog.</b>	<b>AR(2)</b>
21008,650	21029,9	21040,8517
21144,179	21039,3	21052,2388
21206,289	21048,8	21063,6321
21184,039	21058,2	21075,0316
21136,230	21067,7	21086,4373
21173,689	21077,1	21097,8491
21182,529	21086,6	21109,2671
21271,970	21096	21120,6913
21235,669	21105,5	21132,1217
21328,470	21115	21143,5582

Poté byly masivně testovány různé parametry na větším množství dat (různé varianty)

Postup (podrobný postup a použité skripty):

- Do MATLABu byla naimportována data Dow Jonesova indexu (viz Zdroje dat – Yahoo).

První náhodě zvolený den byl s pořadovým číslem 4370 (24.5.2017). Pro odhad volby řádu a množství dat zahrnutých do modelu jsem si vytvořil skript *ChybaAR\_test.m*, který ve spolupráci s dalšími funkcemi dle nastavených parametrů (počet dat a řád AR modelu) postupně počítal predikované hodnoty 1 až 15 kroků dopředu v množství postupně ode dne 1 až po hodnotu, která byla před dnem 4370

Stručný popis algoritmu:

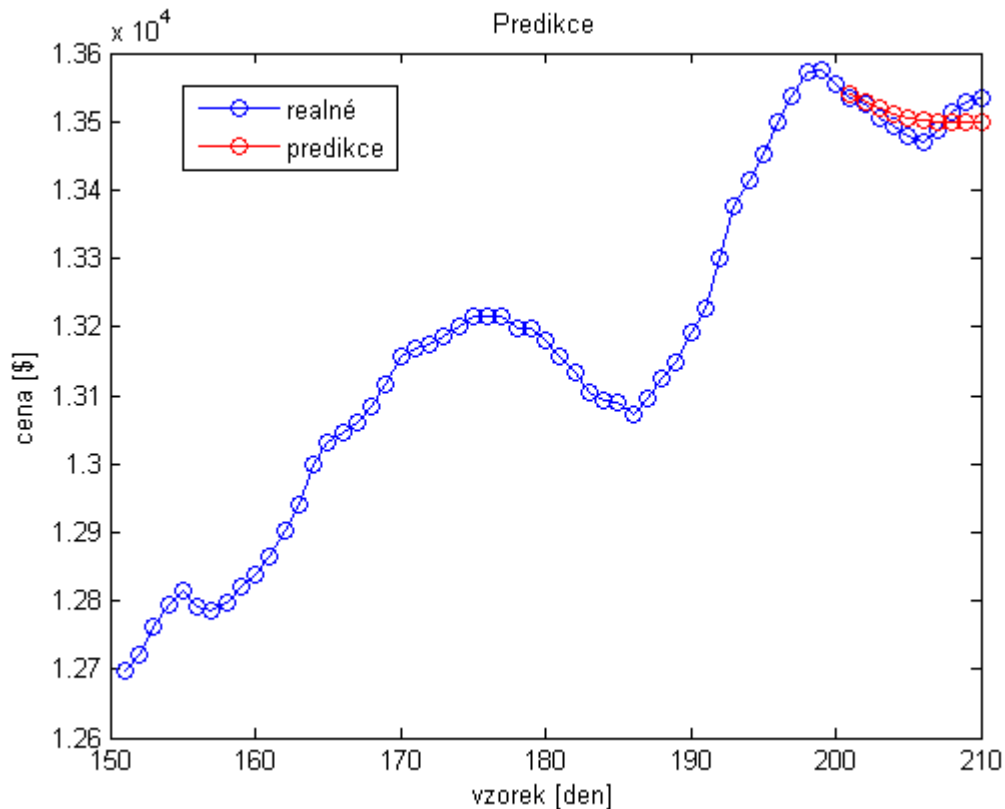
- pro zvolené parametry (např. řád=3 a množství dat zahrnutých do výpočtu modelu (dále uváděno jen jako okno) např. 200 ) se pro vypočítají koeficienty AR-modelu (zde řádu 3) ze závěrečných cen dnů 1 až 199
- na základě výsledných koeficientů se spočítají predikované hodnoty 15 kroků dopředu
- tyto predikované kroky se porovnají se skutečnými, vypočítá se procentuální absolutní odchylka (MAPE) a ta se zapíše
- okno se celé posune o jeden den (tj. další bude den 2 až 200) a opakují se předchozí dva kroky
- celý cyklus jede do doby až se poslední hodnota v okně rovná dnu před dnem 4370 (tj. okno 4170 až 4369)
- se zaznamenaných MAPE se za celý cyklus spočítají průměry pro jednotlivé predikované kroky a ty se včetně parametrů (řád AR modelu a délka okna) zaznamenají do výsledné tabulky.
- tyto cykly se opakují pro všechny zadané parametry

Na základě výsledků z tabulky jsem se dále zkoušel parametry blízké těm které měly úspěšné výsledky. Hodnoty obecně byly velmi špatné a v konečném důsledku i při

jednokrokové predikci horší nežli kdyby byla ponechána hodnota před predikcí

## 6.2 Vyfiltrovaná data pomocí klouzavých průměrů

Po předchozím neúspěchu byla data vyfiltrovan pomocí moving average s různou délkou (maximálně 7).



Obr. 9: Vzorový model DAX po MA7

Chyba modelu je

$$R^2=0,996 \quad (16)$$

Na základě toho byl spuštěno testování pro velké množství dat.

Vybrané údaje jsou v tabulce Tab. 2. Rozdíly po vyfiltrovaní klouzavým průměrem

Na první pohled je vidět, že s predikce s rostoucím vyrovnání se zlepšuje a je poměrně úspěšná pro první krok maximálně třetí

Tab. 2. Rozdíly po vyfiltrování klouzavým průměrem

		<i>Kroků predikce</i>				
		1	2	3	4	5
<i>MA4</i>	<i>Prům. rozdíl predikce</i>	30,6	58,0	78,7	139	170,5
<i>AR(2)</i>	<i>Prům rozdíl poslední</i>	32,6	58,0	78,7	96,2	111,6
<i>MA7</i>	<i>Prům. rozdíl predikce</i>	18,0	33,8	63,3	90,1	118,6
<i>AR(2)</i>	<i>Prům rozdíl poslední</i>	27,6	51,8	73,3	92,9	111,0
<i>MA9</i>	<i>Prům. rozdíl predikce</i>	14,2	30,9	50,7	72,6	96,4
<i>AR2</i>	<i>Prům rozdíl poslední</i>	24,8	47,2	67,7	86,8	104

### 6.3 Regresní model s více vstupy

V článku [1] jeho autor profesor Alaa F. Sheta s kolektivem porovnávají predikci akciového indexu S&P500 pomocí regresního model s více vstupy s predikcí využívající genetického programování. Tento článek a model v něm uvedený byl využit s drobnými úpravami pro další testování.

#### 6.3.1 Použitá data a jejich zdroje

Všechna data jak vstupní, tak výstupní byla vzorkována s po jednotlivých dnech. Graf hodnot S&P500 za celou použitou sadu vypadá následovně:

Vstupní data byla rozdělena do skupin z následujících kategorií:

$S_1$ : Hodnoty zkoumaného aktiva (S&P500 indexu) za předchozí tři období (tj. den-1, den-2, den-3)

$S_2$ : Finanční a ekonomické indikátory (ropa, zlato, CTB3M, AAA)

$S_3$  Hodnota pěti největších společností ze zkoumaného aktiva ( Microsoft, Apple, Amazon, Facebook, Berkshire)

S<sub>4</sub> Kurzy amerického dolaru vůči dalším světovým měnám (japonský jen, britská libra, kanadský dolar)

S<sub>5</sub> Hodnoty dalších významných indexů ( hongkongský HSI, francouzský CAC, britský FTSE a německý DAX)

S<sub>6</sub> Objemy zkoumaného aktiva

S<sub>7</sub> Změny v úrokových sazbách dluhopisů

Data byla získána z pokud možno stejných zdrojů. Již při získávání dat se objevily komplikace.

Některá data se přestala poskytovat na uvedeném zdroji – vyřešeno použitím jiného zdroje, popř. rozdělením dat na část včetně těchto údajů a bez těchto údajů (výnosy korporátních dluhopisů ratingu AAA a BAA).

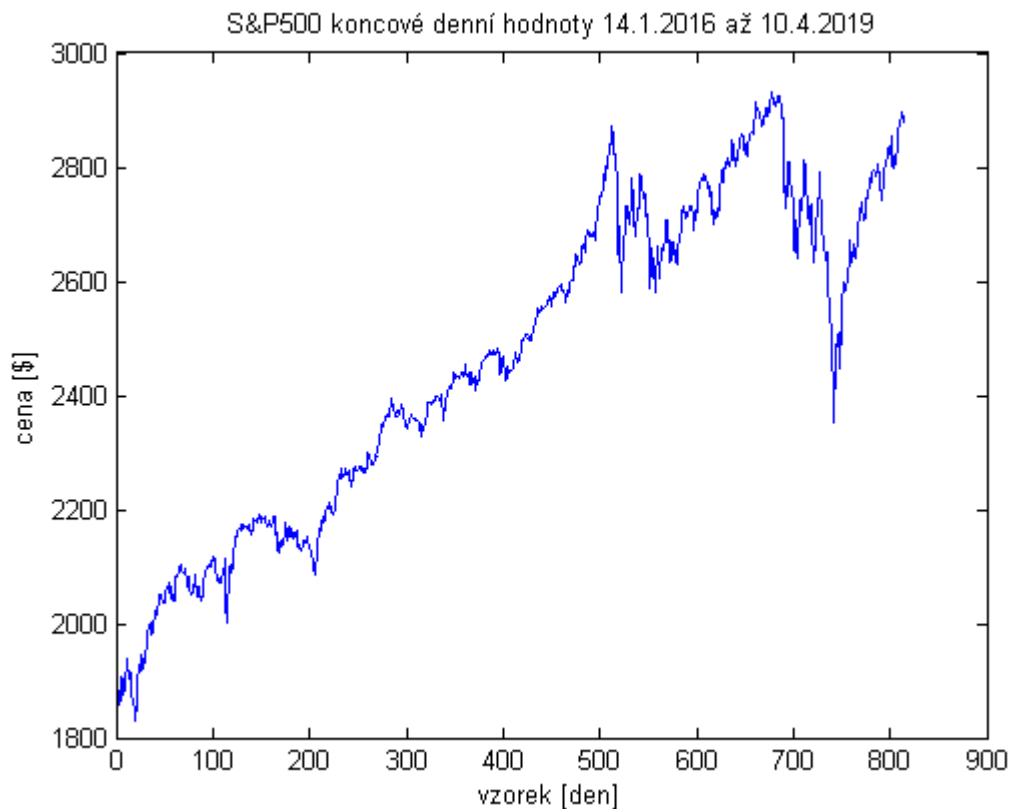
Dále byly nahrazeny akcie s největší vahou za aktuální (článek byl z roku 2015 a váhy jednotlivých akcií se každého čtvrt roku upravují).

Tab. 3. Zdroje pro regresi s více vstupy

<i>vstup</i>	<i>zkratka</i>	<i>popis</i>	<i>zdroj</i>
x <sub>1</sub>	SP(k-1)	Hodnota S&P 500 index den -1	yahoo
x <sub>2</sub>	SP(k-2)	Hodnota S&P 500 index den -2	yahoo
x <sub>3</sub>	SP(k-3)	Hodnota S&P 500 index den -3	yahoo
x <sub>4</sub>	Oil	Relativní změna ceny ropy (crude oil)	Metatrd
x <sub>5</sub>	Gold	Relativní změna ceny zlata	Metatrd
x <sub>6</sub>	CTB3M	Změna tržního výnosu cenných papírů US Treasury při 3měsíční konstantní splatnosti	FRGB
x <sub>7</sub>	AAA	Změna Moodyho výnosu korporátních dluhopisů zařazení Aaa	FRGB
x <sub>8</sub>	MSFT	Hodnota akcie Microsoftu den-1	yahoo
x <sub>9</sub>	AAPL	Hodnota akcie Apple den-1	yahoo
x <sub>10</sub>	AMZN	Hodnota akcie Amazon den-1	yahoo
x <sub>11</sub>	FB	Hodnota akcie Facebook den-1	yahoo
x <sub>12</sub>	BRK-B	Hodnota akcie Berkshire den-1	yahoo
x <sub>13</sub>	USDJPY	Relativní změna kurzu US dolaru a japonského jenu	Metatrd
x <sub>14</sub>	GBPUSD	Relativní změna kurzu US dolaru a britské libry	Metatrd

<i>vstup</i>	<i>zkratka</i>	<i>popis</i>	<i>zdroj</i>
X15	USDCAD	Relativní změna kurzu US dolaru a kanadského dolaru Zdroj: Metatrade	Metatrd
X16	HSI	Hodnota hongkongského Hang Seng indexu den-1	yahoo
X17	FCHI	Hodnota francouzského CAC indexu den-1	yahoo
X18	UKFTSE	Hodnota britského FTSE indexu den-1	yahoo
X19	DAX	Hodnota německého DAX indexu den-1	yahoo
X20	Objem	Relativní změna objemu S&P500	yahoo
X21	CTB6M	Změna tržního výnosu cenných papírů US Treasury 3měsíční konstantní splatnost	FRGB
X22	CTB1Y	Změna tržního výnosu cenných papírů US Treasury 1letý konstantní splatnost	FRGB
X23	CTB5Y	Změna tržního výnosu cenných papírů US Treasury 5ti letý konstantní splatnost	FRGB
X24	CTB10Y	Změna tržního výnosu cenných papírů US Treasury 10ti letý konstantní splatnost	FRGB
X25	BBB	Změna Moodyho výnosu korporátních dluhopisů zařazení Baa	FRGB
X26	DJI	Hodnota Down Jones Industrial Average indexu den-1	yahoo
X27	NASDAQ	Hodnota NASDAQ indexu den-1	yahoo

Zdroj označený yahoo je finance.yahoo.com, Metatrade – software Metatrade, FRGB – H.15 – Federal Reserve Board of Governors



Obr. 10: S&P500 – celá datová sada

### 6.3.2 Příprava dat

Z jednotlivých dat byl vytvořen jeden datový set. V případě neexistence jednotlivých hodnot (výpadky vznikly neobchodováním v daný den, např. svátky v rozdílných zemích) se doplnila hodnota hodnotou předchozí.

Z celkové datové sady byly vytvořeny dvě sady. Jedna včetně výnosů za korporátní dluhopisy Aaa a Baa (od 14.1.2016 do 10.7.2016), druhá bez těchto proměnných (od 14.1.2016 do 10.4.2019) tj. jen 25 vstupních proměnných.

### 6.3.3 Tvorba modelu

Přepíše se rovnice (1) na tvar

$$y(k) = c_1 x_1 + c_2 x_2 + \dots + c_{27} x_{27} + e(k) \quad (17)$$

Kde  $c_1$  až  $c_{27}$  jsou výše uvedené jednotlivé proměnné jdoucí do regresní modelu.

Pro zjištění parametrů  $c_x$  se vytvoří regresní matice  $\hat{X}$ . Po úpravách je výsledná rovnice v maticovém tvaru a poté řešena metodou nejmenších čtverců

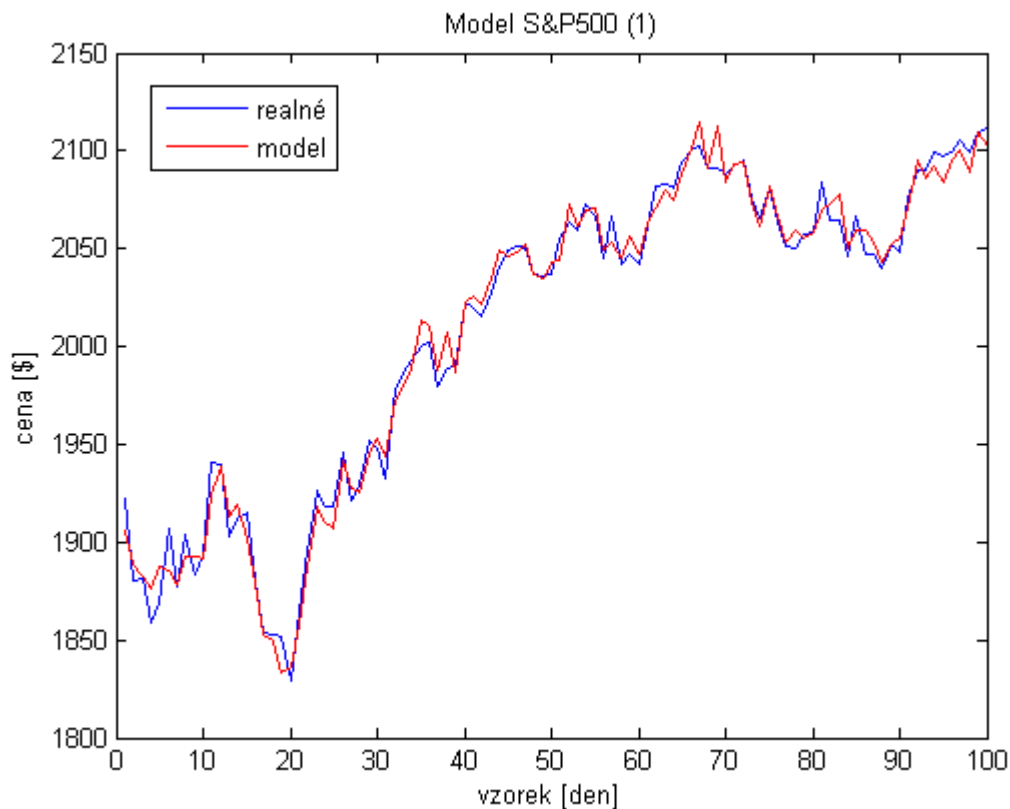
$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_{27}(1) \\ x_1(2) & x_2(2) & \dots & x_{27}(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \dots & x_{27}(N) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{27} \end{bmatrix} + \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix} \quad (18)$$

Po zvolení dat pro modelování od 1 do 100 (odpovídá dnům 14.1.2016 až 8.6.2016) je řešením následující rovnice

$$\begin{aligned} \hat{y} = & 1,2474 \cdot x_1 - 0,0076992 \cdot x_2 + 0,053756 \cdot x_3 + 17,324 \cdot x_4 - 287,49 \cdot x_5 \\ & - 4,2865 \cdot x_6 + 62,971 \cdot x_7 + 3,004 \cdot x_8 + 0,21290 \cdot x_9 + 0,047196 \cdot x_{10} \\ & + 0,42806 \cdot x_{11} + 5,0430 \cdot x_{12} + 293,83 \cdot x_{13} - 198,78 \cdot x_{14} - 890,30 \cdot x_{15} \\ & + 0,0016084 \cdot x_{16} - 0,012273 \cdot x_{17} + 0,0058574 \cdot x_{18} + 0,0026462 \cdot x_{19} \\ & + 3,9600 \cdot x_{20} + 7,6822 \cdot x_{21} + 9,4836 \cdot x_{22} + 39,089 \cdot x_{23} - 225,66 \cdot x_{24} \\ & + 746,31 \cdot x_{25} - 0,082819 \cdot x_{26} - 0,032767 \cdot x_{27} \end{aligned} \quad (19)$$

Model tedy vypadá takto





Obr. 11: S&amp;P500 – Model 1

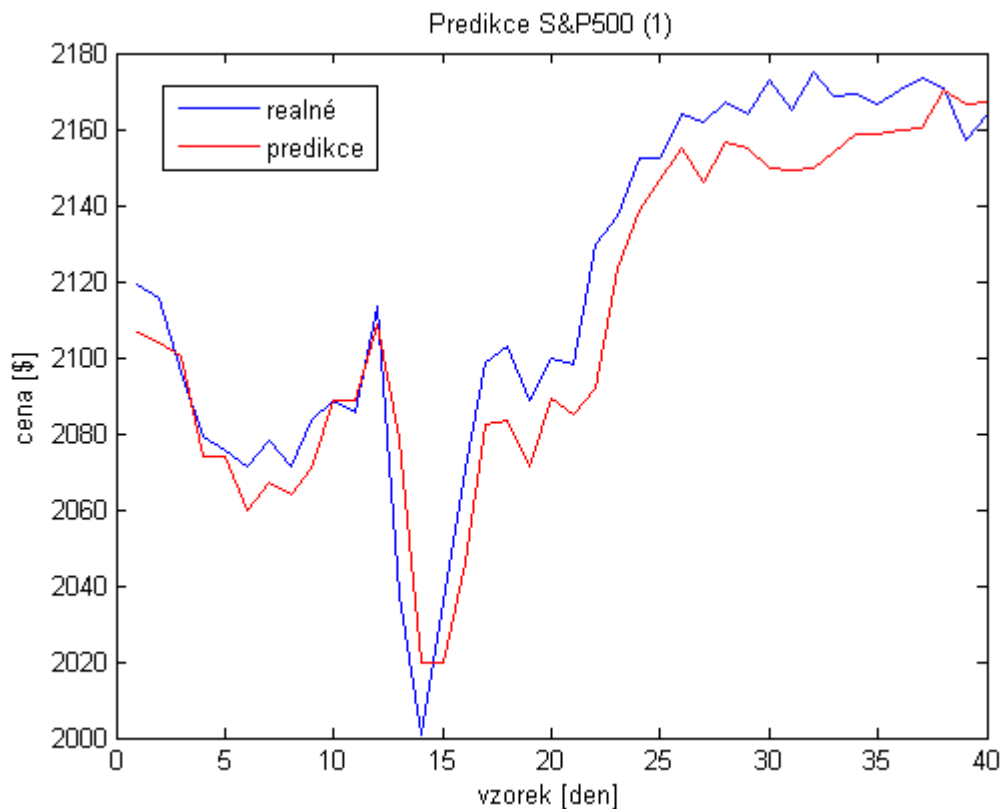
Index spolehlivosti dle vzorce (12) pro tento model je  $R^2=0,989$

Pro zajímavost jakými vahami se se podílely jednotlivé proměnné byly dosazeny střední absolutní hodnoty těchto proměnných do vzorce modelu

$V_1=2503$	$V_2=-15,43$	$V_3=107,6$	$V_4=0,433$	$V_5=-2,411$
$V_6=-0,2076$	$V_7=0,566$	$V_8=157,3$	$V_9=21,3$	$V_{10}=28,65$
$V_{11}=47,29$	$V_{12}=695,4$	$V_{13}=1,792$	$V_{14}=-0,9981$	$V_{15}=-4,875$
$V_{16}=32,32$	$V_{17}=-53,45$	$V_{18}=35,49$	$V_{19}=25,94$	$V_{20}=0,3969$
$V_{21}=0,2669$	$V_{22}=0,2826$	$V_{23}=1,064$	$V_{24}=-4,22$	$V_{25}=4,859$
$V_{26}=-1422$	$V_{27}=-154,4$			

Ukázalo se, že největší váhu mají proměnné  $X_1$  – vlastní předchozí hodnota (den -1),  $X_{26}$  – index Dow Jones a naopak nejmenší  $X_6$  změna výnosů cenných papírů s 3 měsíční splatností

Při použití rovnice (19) pro následujících 40 hodnot (tj. dny 9.6.2016 až 4.8.2016) jsou predikované hodnoty v porovnání se skutečnými v grafu



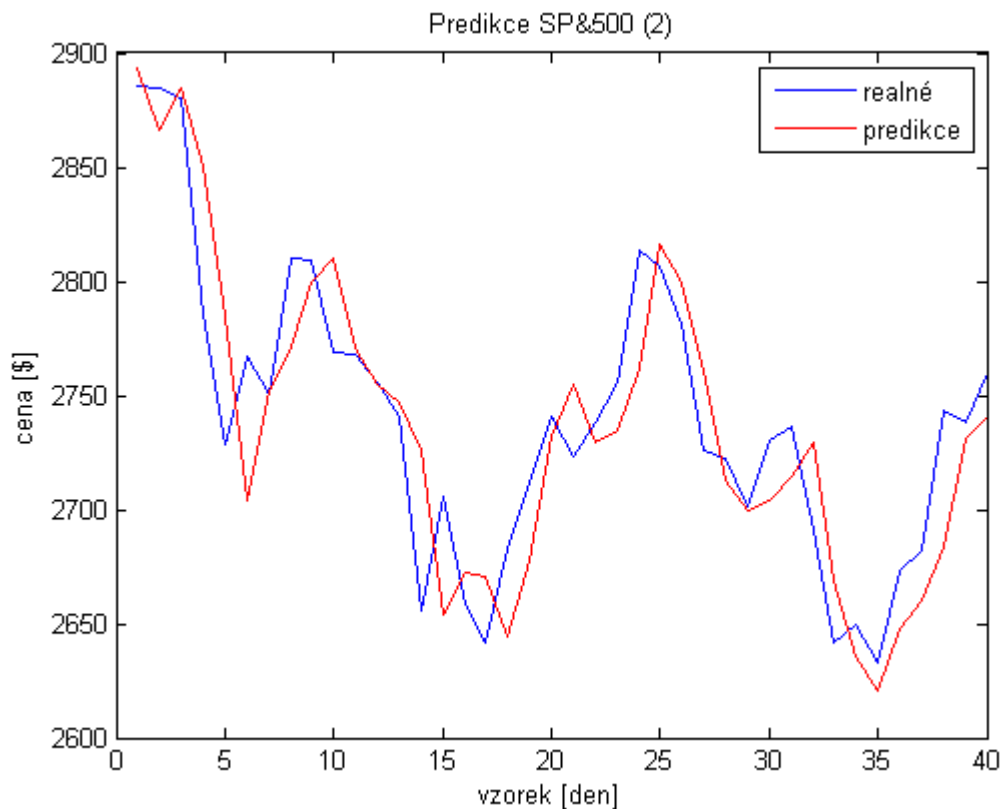
Obr. 12: S&P500 – Predikce 1

Chyba MAPE je 0,6064%

Vše na první pohled poměrně úspěšné, ale pokud si uvědomíme, že:

- SP500 nemění příliš významně, střední absolutní procentní změna za modelovaných 100 hodnot je 0,6085%, což je téměř srovnatelné
- Treningová data jsou počítána
- Hodnoty některých proměnných jsou téměř aktuální (naštěstí s menší váhou, ale asi správnější by bylo také použít den – 1, což jsem také udělal, ale výsledek byl srovnatelný)

Další model jsem udělal úmyslně v místě, kde byla křivka více „zvlněná“ a výsledný graf a chyby vyšly takto:



Obr. 13: S&P500 – Predikce 2

Index spolehlivosti dle vzorce (12) pro tento model je  $R^2=0,9821$

Střední absolutní procentní změna predikce = 0,9347%

## ZÁVĚR

Pro modelování byla získána data a provedeny případné úpravy pro tvorbu modelů. Na vybraných datech byly poměrně úspěšně vytvořeny modely s vysokými indexy spolehlivosti již od nízkých řádů. Dle předpokladu, nepodařilo se najít mezi metodami žádnou velmi úspěšnou metodu a model pro predikci budoucích hodnot vybraných akciových titulů. Autoregresní model je obecně vhodný pro jednokrokovou predikci, v experimentech na burzovních datech bez dalších úprav byl přesto neúspěšný. Toto je zapříčiněno značně stochastickou povahou dat. Přesto po vyfiltrování pomocí klouzavých průměrů metody byly při predikci úspěšnější nežli při úvaze o predikované hodnotě ponechané na poslední hodnotě a relativně úspěšnější v jednokrokové predikci. To ukazuje proč je Moving Average základním kamenem klasických tech-

nických analýz.

Model s více vstupy vypadá na první pohled lépe, ale je si třeba uvědomit, že je to jen jednokroková predikce. Volba parametrů je podstatná, zde se ukázalo, že velkou váhu na výpočet predikované hodnoty měla předchozí hodnota a jeden z hlavních indexů.

Z praktického hlediska při investování je vhodné stále spoléhat na základní statistické ukazatele vývoje akcií u jejich hodnoty využít ve svém řízení rizik ( stop loss – zastavení ztrát). Uvedené metody je možno využít při obchodování v krátkých časových úsecích (tzn. tiky, minuty), což bývá také u vysokofrekvenčního obchodování.[2]

Jako možné vhodné pokračování studie by bylo možné vyhodnocovat pomocí fuzzy logiky (např. obdobně jako makléřské domy doporučují strong buy, buy, hold, sell, strong sell) a zdá se, že při případně větším využití klasických technických indikátorů, popř. hledání formací v grafu (vzorů) např. pomocí neuronové sítě

## SEZNAM POUŽITÉ LITERATURY

- [1] Alaa Sheta [online] [https://www.researchgate.net/profile/Alaa\\_Sheta/publication/280603295\\_Evolving\\_Stock\\_Market\\_Prediction\\_Models\\_Using\\_Multi-gene\\_Symbolic\\_Regression\\_Genetic\\_Programming/links/55bdac9f08aec0e5f4445c8c.pdf](https://www.researchgate.net/profile/Alaa_Sheta/publication/280603295_Evolving_Stock_Market_Prediction_Models_Using_Multi-gene_Symbolic_Regression_Genetic_Programming/links/55bdac9f08aec0e5f4445c8c.pdf). . . , ISSN: 1687-4846
- [2] Cuesta Hektor Analýza dat v praxi. . : computer press, . 978-80-251-4361-2
- [3] DANEL Roman PREDIKCE ČASOVÉ ŘADY POMOCÍ AUTO-REGRESNÍHO MODELU. . 2004, , ,
- [4] DOSTÁL, Petr Soft computing v podnikatelství a veřejné správě. . Brno: CERM, 2015. ISBN 978-80-7204-896-0.
- [5] <https://www.fio.cz/zpravodajstvi/dividendy> Dividendy. . 2019, , ,
- [6] KŘIVÝ Ivan Analýza časových řad. 1.. : Ostravská univerzita v Ostravě, 2012.
- [7] Lixbroker podpora [online] Dostupné z: <https://www.lynxbroker.cz/podpora/pripisovani-dividend/>. . . , , ,

- [8] LITSCHMANNOVÁ Martina Úvod do analýzy časových řad. . . , ,
- [9] MUSÍLEK, Petr. Trhy cenných papírů. . . : Ekopress, 2002. ISBN 80-86119-55-6.
- [10] VIKTORIN, Adam, Analytical Programming With Extended Individuals. . . , , , p. 237-244.
- [11] VOLNÁ Eva Základy softcomputingu. . : Ostravská univerzita v Ostravě, 2012.
- [12] ZELINKA, Ivan, Zuzana KOMÍNKOVÁ OPLATKOVÁ a Roman ŠENKEŘÍK. Aplikace umělé inteligence. . : Univerzita Tomáše Bati ve Zlíně,, 2010. ISBN 978-80-7318-898-6
- [13] ZELINKA, Ivan. Umělá inteligence: neuronové sítě a genetické algoritmy. . . BRNO: VUTIUUM,, 1998. 126 s.ISBN 8021411635.
- [14] Technická analýza [online] dostupné z: <http://www.miras.cz/akcie/technicka-analyza-5.php>. . . , ,
- [15] 15: , Wikipedia [online] Dostupné z [https://cs.wikipedia.org/wiki/Metoda\\_nejmen%C5%A1%C3%Adch\\_%C4%8Dtverc%C5%AF](https://cs.wikipedia.org/wiki/Metoda_nejmen%C5%A1%C3%Adch_%C4%8Dtverc%C5%AF) ,
- [16] Zákon o obchodních společnostech a družstvech. . : , 2017. §256 90/2012 Sb
- [17] [ONLINE] <https://www.etrading.sk/cz/technicka-analyza/>. . . , , ,

**SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK**

ČNB	Česká národní banka
USD	US dolar – americký dolar.
NYSE	New York Stock Exchange, od roku 2017 po fúzi s panevropským trhem Eronext správně New York Stock Exchange Euronext (New Yorská obchodní burza Euro-next).
P/E	Poměr tržní ceny a zisku na akcii (z anglického price to earnings
AR	Autoregresní (model)
MA	Moving average (klouzavý průměr)
Atd.	A tak dále
Popř.	Popřípadě
S&P50	Standard & Poor's 500 (index)

0

**SEZNAM OBRÁZKŮ**

Obr. 1: Zobrazení jen koncových kurzů.....	14
Obr. 2: Sloupcový graf.....	15
Obr. 3: Svíčkový graf.....	15
Obr. 4: Dow Jones Industrial Average – celá datová sada.....	30
Obr. 5: Vzorový AR model DJI denní vzorkování.....	31
Obr. 6: Vzorový AR model DJI predikce komplet.....	32
Obr. 7: Vzorový AR model DJI detail predikce.....	32
Obr. 8: Vzorový model DJI detail genetické programování.....	33
Obr. 9: Vzorový model DAX po MA7.....	35
Obr. 10: S&P500 – celá datová sada.....	39
Obr. 11: S&P500 – Model 1.....	41
Obr. 12: S&P500 – Predikce 1.....	42
Obr. 13: S&P500 – Predikce 2.....	43

## SEZNAM TABULEK

Tab. 1. Výsledky predikce vzorový DJI.....	32
Tab. 2. Zdroje pro regresi s více vstupy.....	35

## SEZNAM PŘÍLOH

Příloha P 1: Název přílohy.



