

**EKONOMICKÁ UNIVERZITA V BRATISLAVE
NARODOHOSPODÁRSKA FAKULTA**

Evidenčné číslo: 101006/B/2024/36146475541352196

PREDIKCIA FINANČNÝCH ČASOVÝCH RADOV

Bakalárska práca

2024

Pavína Hoffmannová

**EKONOMICKÁ UNIVERZITA V BRATISLAVE
NARODOHOSPODÁRSKA FAKULTA**

PREDIKCIA FINANČNÝCH ČASOVÝCH RADOV

Bakalárska práca

Študijný program:	Financie, bankovníctvo a poisťovníctvo
Študijný odbor:	Financie, bankovníctvo a poisťovníctvo
Školiace pracovisko:	Katedra financií
Vedúci záverečnej práce:	Mgr. Marek Káčer, PhD.

2024

Pavína Hoffmannová

Pod'akovanie

Touto cestou by som sa chcela poďakovať môjmu školiteľovi Mgr. Marekovi Káčerovi, PhD., za jeho odborný prístup, cenné rady a odporúčania, ktoré mi pomohli pri vypracovaní mojej bakalárskej práce.

ABSTRAKT

HOFFMANNOVÁ, Pavlína: *Predikcia finančných časových radov*. - Ekonomická univerzita v Bratislave. Národohospodárska fakulta: Katedra financií. Vedúci záverečnej práce: Mgr. Marek Káčer, PhD. – Bratislava: NHF EU, 2024, 48 str.

Cieľom bakalárskej práce bolo modelovanie vybraných časových radov a ich predikcia. Bakalársku prácu sme rozdelili do troch kapitol. Obsahuje 33 grafov a 7 tabuliek. V prvej kapitole sme sa oboznámili s časovými radmi, definovali sme si ich. Dozvedeli sme sa, aké poznáme rozdelenie časových radov, oboznámili sme sa s jednotlivými zložkami, ktoré tvoria časový rad. Pozreli sme sa na analýzu časových radov, charakterizovali sme si finančné časové rady a povedali sme si niečo aj o ich vlastnostiach. V ďalšej časti sme sa venovali predikcii a postupu pri predikovaní časových radov. Dozvedeli sme sa aké metódy sa používajú pri predikcii časových radov a tieto metódy si bližšie popísali. Druhá kapitola bola venovaná cieľom a metodike bakalárskej práce. Oboznámili sme sa v nej s akým programom budeme pri predikcii časových radov pracovať, a popísali sme si aké modely budeme v našej práci aplikovať pri predikcii časových radov. Tretia kapitola bola zameraná a venovaná samotnému porovnaniu dvoch vybraných časových radov, ktoré sme si charakterizovali, modelovali a predikovali a v neposlednom rade porovnali a vyhodnotili.

Kľúčové slová:

časové rady, finančné časové rady, modely ARIMA

ABSTRAKT

HOFFMANNOVÁ, Pavlína: *Financial time series prediction*. – University of Economics in Bratislava. Faculty of National Economy. Department of Finance. – Leader of the final thesis: Mgr. Marek Káčer, PhD. – Bratislava: NHF EU, 2024, 48 p

The aim of the bachelor thesis was to model time series and their prediction. We divided the work into three chapters. It contains 33 graphs and 7 tables. The first chapter is devoted to the description of time series in general. Their features, basic characteristics, types and analysis. We looked at the analysis of time series, we characterized the financial time series and we said something about their properties. In the next part, we focused on the prediction and the procedure for predicting time series. We learned what methods are used in predicting time series and described these methods in more detail. The second chapter was devoted to the aim and methodology of the bachelor thesis. We got acquainted with the program we will work with in predicting time series, and we described what models we will run in our work in predicting time series. The third chapter was focused and devoted to the comparison of two selected time series, which we characterized, modeled and predicted and last but not least, compared and evaluated.

Key words:

time series, financial time series, ARIMA models

Obsah

Obsah.....	7
Úvod.....	8
1 Riešenie problematiky súčasného stavu doma a v zahraničí	9
1.1 Časové rady	9
1.1.1 Definícia	9
1.1.2 Rozdelenie časových radov	9
1.1.3 Klasický (formálny) rozklad	10
1.1.4 Analýza časových radov	13
1.2 Finančné časové rady	14
1.2.1 Charakteristika	14
1.2.2 Vlastnosti finančných časových radov	14
1.3 Predikcia.....	15
1.3.1 Postup pri predikovaní.....	16
1.4 Metódy	18
1.4.1 Autoregresia (AR).....	18
1.4.2 Proces klzavých priemerov (MA)	19
1.4.3 Box-Jenkinsové modely	20
1.4.4 Exponenciálne vyrovnávanie	21
2 Cieľ a metodika práce	23
3 Výsledky práce a diskusia	24
3.1 Analýza dát časového radu maloobchodného predaja v USA.....	24
3.1.1 Základná charakteristika dát	25
3.1.2 Modelovanie dát z časového radu	26
3.1.3 Tréningové vzorky a ich predikcia.....	31
3.2 Analýza dát časového radu HDP v Spojených štátoch amerických	32
3.2.1 Základná charakteristika dát	33
3.2.2 Modelovanie dát z časového radu	33
3.2.3 Tréningové vzorky a ich predikcia.....	38
3.3 Hodnotenie výsledkov.....	40
Záver.....	41
Zoznam použitej literatúry:.....	42
Zoznam ilustrácií.....	44
Zoznam tabuliek.....	44
Prílohy	45

Úvod

S časovými radmi, ako hodnotami určitej veličiny zoradenej do chronologického postupu sa stretneme v takmer všetkých oblastiach ľudského života. Jedna z najbežnejších oblastí, kde sa časové rady vyskytujú, je ekonómia. Dôležitosť časových radov pre dnešný svet tkvie v dôkladnej analýze minulých a súčasných dát, ktoré vedú k zodpovednému rozhodovaniu. Žijeme v období, kedy nie je možné robiť dôležité ekonomické rozhodnutia bez prepracovanej a dôkladnej analýzy základných ekonomických ukazovateľov a ich vzťahov.

V bakalárskej práci si popíšeme časové rady, ktoré môžeme nájsť okolo nás. Konkrétne sa budeme zaoberať ekonomickými a predovšetkým finančnými časovými radmi, v ktorých sa používajú makroekonomické ukazovatele. V práci popisuje tieto problematiky ako teoreticky, tak i prakticky.

Bakalársku prácu sme si rozdelili do troch kapitol. V prvej kapitole sa oboznámime s časovými radmi, definujeme si ich. Dozvieme sa, aké poznáme rozdelenie časových radov, oboznámime sa s jednotlivými zložkami, ktoré tvoria časový rad. Pozrieme sa na analýzu časových radov, charakterizujeme si finančné časové rady a povieme si niečo aj o ich vlastnostiach. V ďalšej časti bakalárskej práce sa budeme venovať predikcii a postupu pri predikovaní časových radov. Dozvieme sa, aké metódy sa používajú pri predikcii časových radov a tieto metódy si bližšie popíšeme.

Druhá kapitola bakalárskej práci je venovaná cieľom a metodike bakalárskej práce. Oboznámime sa v nej s akým programom budeme pri predikcii časových radov pracovať, a popíšeme si aké modely budeme v našej práci aplikovať pri predikcii časových radov.

Tretia kapitola bakalárskej práce sa venuje samotnej praktickej časti v ktorej si porovnáme dva vybrané časové rady, ktoré si budeme charakterizovať, modelovať a predikovať a v neposlednom rade si ich vyhodnotíme a porovnáme.

Hlavným cieľom mojej bakalárskej práce je priblíženie, analyzovanie a modelovanie časových radov. Ako môžeme vhodne zvolenými metódami ľahko a presne predikovať časové rady a simulovať vplyv pôsobiacej zmeny vo vývoji finančnej časovej rady. Ďalším cieľom bakalárskej práce je využitie získaných poznatkov pri predikcii budúceho chovania finančnej časovej rady. Postupy, ktoré použijeme, môžu byť založené na princípe opakovania sa minulosti. Tento predpoklad môžeme v praxi vykazovať s rôznymi presnosťami a preto je úlohou bakalárskej práce zistiť vhodné postupy pri predpovedaní časovej rady.

1 Riešenie problematiky súčasného stavu doma a v zahraničí

1.1 Časové rady

Pokiaľ stojíme pred problémom analýzy určitého ekonomického javu, určite sa stretneme s pojmom časové rady.

1.1.1 Definícia

Dáta, ktoré potrebujeme analyzovať rôznymi štatistickými metódami, sú väčšinou vo forme tzv. časových radov. Môžeme ich nájsť alebo získať v mnohých oblastiach od ekonómie, štatistiky, predpovedaní počasia až po strojárstvo, medicínu alebo astronómiu.

Pod pojmom časové rady si môžeme predstaviť chronologické usporiadanie sledovaných hodnôt náhodnej veličiny. Časové rady môžu byť napríklad indexy spotrebiteľských cien za určité časové obdobie, počet prijatých žiakov na základnú školu, počet narodených detí na Slovensku, ceny akcií na burze, údaje o dennej návštevnosti kultúrnych pamiatok, priemerné mzdy zamestnancov a pod. Dáta v tvare časových radov alebo časovo pozbierané dáta nájdeme napríklad vo vede, technike, ekonómii, priemysle ale i v mnohých ďalších oblastiach všade okolo nás.

1.1.2 Rozdelenie časových radov

Pre správny výber metódy daného ukazovateľa je potrebné prihliadať na charakter a vlastnosť jeho hodnôt. Preto rozlišujeme viacero druhov časových radov.

1. Na základe charakteru dát:

- a. intervalové – sú dáta vyberané od začiatku do konca určitého intervalu, závisí od daného časového obdobia;
- b. okamihové – dáta vzťahujúce sa na okamih, nezávisia od dĺžky časového intervalu.¹

¹ HINDLS, Richard - HRONOVÁ, Stanislava - SEGER Jan. *Statistika pro ekonomy*. vyd. 7. Praha: Profesional Publishing, 2002. 247 s. ISBN 80-864-1930-4.

2. Na základe periodicity zisťovania hodnôt:
 - a. krátkodobé – sledovanie hodnôt za obdobie, ktoré je kratšie ako rok, sledovanie hodnôt štvrťročne, mesačne, denne, polročne;
 - b. dlhodobé – sledovanie hodnôt za rok alebo dlhšie časové obdobie, dá sa pri nich lepšie vystihnúť cyklické kolísanie.
3. Na základe druhu údajov:
 - a. extenzitné – určujú sa v naturálnych jednotkách alebo menových jednotkách, určujú veľkosť alebo množstvo;
 - b. intenzitné – určujú stav alebo intenzitu výskytu v tvare podielu, v dvoch meracích jednotkách alebo v percentách.²
4. Na základe spôsobu vyjadrenia údajov:
 - a. naturálne – hodnoty ukazovateľa sú vyjadrované v naturálnych jednotkách;
 - b. peňažné – hodnoty ukazovateľa sú vyjadrované v peňažných jednotkách.

S týmto základným členením súvisí predovšetkým tvar ekonomických časových radov, napr. čím je interval sledovania dlhší, tým sú časové rady vyhladenejšie. Dôležitou časťou je aj poradie časových radov, spôsob ako na seba jednotlivé hodnoty v časových radoch nadväzujú. Poradie určuje ich tvar a charakteristické vlastnosti.³

1.1.3 Klasický (formálny) rozklad

Vo všeobecnosti vieme časový rad rozdeliť na štyri zložky, ktorými sú: trend, sezónna zložka, cyklická zložka a reziduálna zložka.

² SODOMOVÁ, Eva a kol. *Štatistika pre bakalárov*. 2. vyd. Bratislava : EKONÓM, 2013. s. 174. ISBN 978-80-225-3614-1.

³ ARLT, Jozef- ARLTOVÁ, Markéta. *Ekonomické časové rady*. vyd. 1. Praha: Profesional Publishing a.s., 2009. 14 s. ISBN 978-80-8694-685-6.

1.1.3.1 Trend

Existuje, keď dochádza k dlhodobému nárastu alebo poklesu. Niekedy ho budeme označovať ako „meniaci sa smer“, kedy môže prejsť z rastúceho trendu na klesajúci trend. Ide o dlhodobé správanie sa časových radov a vzniká ak naňho pôsobia sily / sila konštantne, čiže v tom istom smere. Trend môžeme v časových radoch popísať trendovými funkciami, kľzavými priermi alebo pomocou kľzavých mediánov. Trendové funkcie sa používajú ak časovému radu zodpovedá určitá funkcia času, a to napr. lineárna, kvadratická alebo exponenciálna. Kľzavý priemer alebo kľzavý medián sa pri modelovaní trendu používa, ak sa časový rad v dôsledku vplyvu nesystematickej zložky, ktorá je veľmi silná, vyvíja nerovnomerne alebo vykazuje extrémne hodnoty.⁴

Najčastejšie používané typy trendu sú:

Konštantný trend: $T_t = \beta_0, t = 1, \dots, n.$

Lineárny trend: $T_t = \beta_0 + \beta_1 t, t = 1, \dots, n.$ kde koeficienty β_0, β_1 odhadneme použitím lineárnej regresie.

Kvadratický trend: $T_t = \beta_0 + \beta_1 t + \beta_2 t^2, t = 1, \dots, n.$ Koeficienty $\beta_0, \beta_1, \beta_2$ odhadneme pomocou lineárnej regresie.

Exponenciálny trend: $T_t = \alpha \times \beta^t, t = 1, \dots, n, \beta > 0.$ ⁵

1.1.3.2 Sezónna zložka

Sezónny vzorec nastáva, keď je časový rad ovplyvnený sezónnymi faktormi, ako je čas v roku alebo deň v týždni. Sezónnosť má vždy pevnú a známu frekvenciu. Medzi hlavné faktory, ktoré ovplyvňujú vznik sezónnych výkyvov patria klimatické podmienky ako i ľudské zvyky. Sezónnu zložku ľahko odhalíme pomocou korelogramu a vyžaduje si elimináciu pri postupe predikcie časového radu, ktorý sledujeme.

⁴ HYNDMAN, J Rob. – ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 2. vydanie. [elektronický zdroj]. 2. vydanie Melbourne : OTexts. 2018.. ISBN 978-0987507112. [cit. 23.10.2021]. Dostupné na: <https://otexts.com/fpp2>.

⁵ ARLT, Josef – ARLTOVÁ Markéta – RUBLÍKOVÁ, Eva. 2002. *Analýza ekonomických časových řad a příklady*. 1. vydanie. Praha: Vysoká škola ekonomická Praha, 2002. s. 148. ISBN 80-245-0307-7.

1.1.3.3 Cyklická zložka

Cyklus nastáva, keď údaje vykazujú vzostupy a poklesy, ktoré nemajú pevnú frekvenciu. Tieto výkyvy sú zvyčajne spôsobené ekonomickými podmienkami a často súvisia s „obchodným cyklom.“ Trvanie týchto výkyvov, na rozdiel od sezónnej zložky, je zvyčajne najmenej 2 roky.⁶

1.1.3.4 Reziduálna zložka

Považujeme ju za zložku, ktorá sa dá najmenej štatisticky definovať. Radíme sem nesystematické výkyvy, ktoré nemajú sezónny a ani cyklický charakter. Je tvorená náhodnými pohybmi (fluktuáciami) v priebehu časového radu, ktoré nemajú rozpoznateľný – systematický charakter. Nepočíta sa preto medzi tzv. systematické zložky časového radu. Pokrýva chyby v meraní údajov, ktoré tvoria časový rad. Predpokladáme, že reziduálna zložka je biely šum, čiže nekorelované náhodné veličiny, vyznačujúce sa nulovou strednou hodnotou.⁷

Nie vždy musí časový rad obsahovať všetky štyri zložky. Rozlišuje dva typy modelov, ktoré sú používané pri časových radoch, ide o aditívne a multiplikačné modely.

Pri aditívnom modeli predpokladáme, že zložky, ktoré pôsobia na časové rady sú navzájom nezávislé. Sezónne výkyvy sa určujú prostredníctvom sezónnych priemerov. Čím je väčšia hodnota, tým väčšie sú sezónne výkyvy. Komponenty multiplikačného modelu nie sú nezávislé a môžu sa vzájomne ovplyvňovať. Sezónne výkyvy sú vyjadrené prostredníctvom sezónnych indexov. Najvhodnejší výber daného modelu závisí od veľkosti výkyvu sledovaných údajov od trendu. Multiplikačný model použijeme, ak sa výkyvy často krát zväčšujú alebo zmenšujú. Naopak, ak sú výkyvy približne rovnakej hodnoty, použijeme aditívny model.

Aditívny model: $Y_t = T_t + S_t + C_t + N_t$

Multiplikačný model: $Y_t = T_t \times S_t \times C_t \times N_t$ ⁸

⁶ HYNDMAN, J. Rob – ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 2. vydanie. [elektronický zdroj]. 2. vydanie Melbourne : OTexts. 2018.. ISBN 978-0987507112. [cit. 23.10.2021]. Dostupné na: <https://otexts.com/fpp2>.

⁷ CIPRA, Tomáš. *Finanční a pojistné vzorce*. vyd. 1 Praha: Grada Publishing a.s., 2006. 323-324 s. ISBN 80-247-1633-X.

⁸ ANDERSON, David R. et al. *Statistics for Business and Economics China*: RR Donnelley. 2015. 512-516 s. ISBN 978-1-4080-72223-3

1.1.4 Analýza časových radov

Pri analyzovaní časových radov je našim cieľom konštrukcia modelu, ktorý najlepšie popisuje, aké správanie môže od skúmaných časových radov očakávať. Pomocou vytvoreného modelu pochopíme správanie a mechanizmus časového radu. Časový rad môže vykazovať pravidelné cykly, trend a pod. Pomocou tejto konštrukcie môžeme tiež zistiť, čím sú hodnoty našej časovej rady ovplyvňované a naopak, na čom sú pravdepodobne nezávislé. S ich pomocou zistíme, čo nám pomôže čiastočne ovládať a optimalizovať správanie časovej rady, či je to výber a voľba vhodných štatistických testov.

Analýza časových radov, vrátane predpovedaní jej budúceho správania, sa stáva jednou z najdôležitejších oblastí v rozvoji súčasnej štatistiky. Hlavným dôvodom rastúceho významu tejto disciplíny je fakt, že sa úspešne vyrovnáva s popisom dynamických systémov, s ktorými často prichádzame do kontaktu.⁹

Voľba vhodnej metódy k analyzovaniu konkrétnej časovej rady závisí na niekoľkých faktoroch. Jedným z najdôležitejších je účel analýzy, kde je dôležitým rozhodnutím je, aký bude cieľ našej analýzy. Ďalším faktorom je typ časovej rady. Typom alebo charakterom časovej rady môžeme rozumieť dĺžku časovej rady, ale aj predpokladané rozdelenie náhodných veličín, ktoré tvoria časovú radu a ich vzájomný vzťah. V poslednom rade je dôležitá i skúsenosť a znalosť analytika, ktorý rozbor časovej rady sprevádza a jeho počítačové a programové vybavenie.¹⁰

⁹ CIPRA, Tomáš. *Finanční a pojistné vzorce*. vyd. 1 Praha: Grada Publishing a.s., 2006. 323-324 s. ISBN 80-247-1633-X.

¹⁰ CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii: celostátní vysokoškolská učebnice pro stud. Matematicko - fyz. fakult studijních odb. 11 Fyzik. matematické vědy*. Praha: Státní nakladatelství technické literatury, 1986. 126 s.

1.2 Finančné časové rady

Finančné časové rady môžeme chápať ako radu hodnôt ekonomického ukazovateľa, ktorá je usporiadaná v čase od minulosti po prítomnosť.

1.2.1 Charakteristika

Tak ako sme uviedli pri definícii časového radu, že časový rad je postupnosť hodnôt sledovanej veličiny v závislosti od času, tak aj finančné časové rady predstavujú časovú postupnosť sledovaných údajov. Je to kvôli tomu, že každodenne porovnávame vývoj cien na svetových trhoch, výmenné kurzy alebo akciové indexy. Pri analýze finančných časových radov musíme rátať s prvkom neistoty, vzhľadom na to, že sa tu objavuje kolísanie aktív. Kvôli tomu je veľmi dôležité, aká metóda a teória je vybraná pri analýze finančných časových radov.

1.2.2 Vlastnosti finančných časových radov

Aj pri finančných časových radoch platí, že sa jedná o chronologicky usporiadanú postupnosť, tak ako pri časových radoch, podobných údajov. Finančné ako i ekonomické časové rady ukazujú postupnosť sledovaných údajov v čase. Je to z dôsledku denného sledovania vývoja cien, akciových indexov, výmenných kurzov. Informácie o cenách sú dôležité pre budúcich investorov a na základe nich sa rozhodujú do budúcnosti a preto sú zaznamenávané s vysokou frekvenciou. Označenie finančných časových radov je spojené s definíciou a fungovaním finančných trhov. „Miesto kde sa stretáva ponuka voľných finančných prostriedkov a dopyt po týchto prostriedkoch, ktoré sa využívajú ako investície nazývame finančný trh. Ponuka predstavuje prebytkové subjekty, ktoré majú voľné zdroje a chcú ich umiestniť na finančnom trhu. Hovoríme o investoroch alebo veriteľoch. Dopyt predstavuje deficitné subjekty, ktoré nemajú dostatok finančných zdrojov. Hovoríme o dlžníkoch alebo ide o emitentoch cenných papierov, ktorí takto získavajú finančné zdroje. Po boku veriteľov a dlžníkov vystupujú rozličné ekonomické subjekty ako napríklad domácnosti, vlády alebo podniky.“¹¹

¹¹ SIVÁK, Rudolf a kol. *Financie*. Bratislava: Wolters Kluwer, 2019 34 s. ISBN 978-80-7598-533-0.

Peňažný, kapitálový, devízový, komoditný a poistný trh je súčasť finančného trhu. Cenný papier je jadro finančného trhu a jeho hlavnou informáciou je cena. „Ceny sú pozorované v určitej časovej frekvencii a tvoria tak časové rady. Časové rady, rady vyplývajúce z cien alebo charakterizujúce ceny a ich vývoj, sa označujú ako finančné časové rady.“¹²

Finančné časové rady sú charakterizované ako časové rady odvodené z cien alebo charakterizujúce ceny. V prípade finančných časových radov sa vychádza z predpokladu, že cena nemôže byť záporné číslo, predpokladá sa teda, že by hodnoty týchto časových radov mohli byť generované logaritmicko-normálnym rozdelením.¹³

1.3 Predikcia

Ekonomické ako i finančné procesy môžeme zaradiť k tým najzložitejším, a to kvôli tomu, že spoločnosť dlhodobo vytvára javy, vyznačujúce sa chaotickým správaním. V takýchto prípadoch je zložitejšie odhadnúť budúci vývoj. V súčasnosti sa vo veľkej miere aj za pomoci výpočtovej techniky pristúpilo k využívaniu a používaniu pokročilých metód, ktoré nám poskytujú lepšie a rýchlejšie výsledky. K týmto metódam patria napríklad neurónové siete, genetické algoritmy ako i fuzzy logika. Môžeme ich použiť ako samostatné metódy, ale dajú sa použiť aj tak, že ich vzájomne skombinujeme.

Predikcia patrí medzi procesy, ktoré začínajú pred udalosťou - výpočtom predikcie, prebiehajú pri procese – korekcia predikcie a končia po udalosti – vyhodnotenie predikcie. Dôležitú úlohu pri predikčnom modeli má vyhodnotenie jeho kvality, ktoré prevádzame za predpokladu skutočných a predpovedaných znalostí hodnôt. Na posúdenie výkonnosti najčastejšie používame tieto kritériá, ktoré vyjadrujeme v percentách. Ide o :

1. Strednú kvadratickú chybu MSE (Mean squared error)

$$MSE = \frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2$$

¹² ARLT, Josef-ARLTOVÁ, Markéta. *Finančný časové rady : vlastnosti, metódy modelování, příklady a aplikace*. Praha: Grada Publishing a.s., 2003. 12 s. ISBN 80-247-0330-0

¹³ CIPRA, Tomáš. *Finanční a pojistné vzorce*. vyd. 1. Praha: Grada Publishing a.s., 2006. 336 s. ISBN 80-247-1633-X.

2. Strednú absolútnu chybu MAE (Mean absolute error)

$$MAE = \frac{1}{N} \sum_{t=1}^N |x_t - \hat{x}_t|$$

3. Strednú relatívnu chybu MAPE (Mean absolute percentage error)

$$MAE = \frac{100}{N} \sum_{t=1}^N \left| \frac{x_t - \hat{x}_t}{x_t} \right|$$

1.3.1 Postup pri predikovaní

Bez ohľadu na to, aký sme zvolili prístup pri predikcií, máme kroky, ktoré musíme dodržiavať. Najčastejšie to je päť krokov, a to pri akomkoľvek modeli.

1. Definícia problému

Patrí často medzi najťažšiu časť prognózovania. Dôkladné definovanie problému vyžaduje pochopenie spôsobu, akým sa budú prognózy používať, kto ich vyžaduje a ako funkcia prognózovania zapadá do organizácie vyžadujúcej prognózy. Analytik musí venovať potrebný čas rozhovormi s každým, kto sa bude podieľať na zhromažďovaní údajov, údržbe databáz a využívaním predpovedí na budúce plánovanie.

1. Zhromažďovanie informácií

Vždy sú potrebné aspoň dva druhy informácií, a to: štatistické údaje a nahromadené odborné znalosti ľudí, ktorí zbierajú údaje a používajú predpovede. Často je ale ťažké získať dostatok historických údajov, aby bolo možné, prispôbiť sa dobrému štatistickému modelu a v takom prípade možno použiť metódy úsudkového prognózovania.

2. Predbežná analýza

V tejto fáze vykonávame analýzu údajov a informuje sa, či sú všetky zhromaždené údaje prospešné alebo nie. Je tiež užitočná pri odhaľovaní vzorov alebo trendov a tým pádom nám pomáha pri výbere modelov, ktoré sa najlepšie hodia. Robí sa tu dôležitá kontrola nepotrebných, nadbytočných údajov. Zmyslom predbežnej analýzy je prevedenie prvotnej selekcie a uľahčenie, tak orientácie v posudzovanej oblasti.

3. Výber a montáž modelov

V tomto kroku vyberáme najlepší model, ktorý sa má použiť. Závisí to od dostupnosti historických údajov, sily vzťahov medzi premennou prognózy a akýmikoľvek vysvetľujúcimi premennými a od spôsobu, akým sa majú prognózy používať. Zvyčajne sa porovnávajú dva až tri potencionálne modely.¹⁴

4. Kontrola platnosti modelu

5. Použitie a vyhodnotenie prognostického modelu

Po výbere modelu a dohodou jeho parametrov sa model použije na tvorbu prognóz. Výkonnosť modelu je možné správne vyhodnotiť až po sprístupnení údajov za prognózované obdobie.

6. Monitorovanie výkonnosti modelu prognózy

Po sprístupnení údajov vieme vyhodnotiť výkonnosť modelu za prognózové obdobie. Je vyvinuté veľké množstvo metód, ktoré nám vedia pomôcť pri posudzovaní presnosti prognóz.¹⁵

¹⁴ HYNDMAN, J. Rob. – ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 2. vydanie. [elektronický zdroj]. 2. vydanie Melbourne : OTexts. 2018.. ISBN 978-0987507112. [cit. 23.10.2021]. Dostupné na: <https://otexts.com/fpp2>.

¹⁵ HYNDMAN, J. Rob. – ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 2. vydanie. [elektronický zdroj]. 2. vydanie Melbourne : OTexts. 2018.. ISBN 978-0987507112. [cit. 23.10.2021]. Dostupné na: <https://otexts.com/fpp2>.

1.4 Metódy

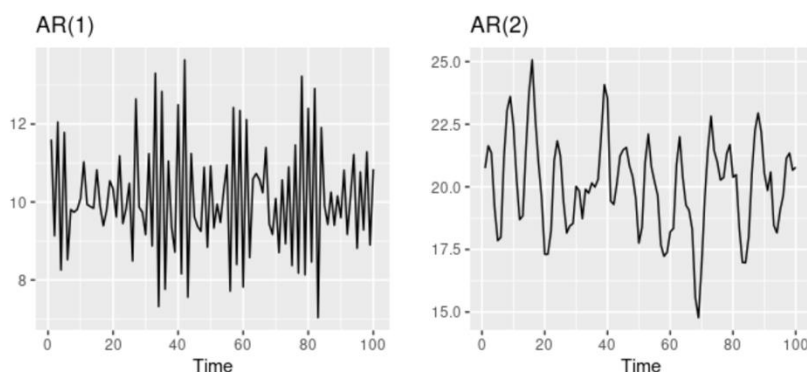
Predikčný model má mimoriadny význam, keďže odhaľuje štruktúru časových radov. Modely časových radov sú buď lineárne alebo nelineárne, je to na základe toho, či je súčasná hodnota série buď lineárnou alebo nelineárnou funkciou predchádzajúcich pozorovaní. Poznáme dva najčastejšie používané lineárne modely auto-regresné modely (AR) a modely kľzavého priemeru (MA). Výsledok týchto dvoch modelov je vznik nového modelu označeného ako Auto-regresný kľzavý priemer (ARMA) a podobný autoregresný integrovaný model kľzavého priemeru (ARIMA).

1.4.1 Autoregresia (AR)

Tento typ modelu môžeme chápať ako vstupný krok pre aplikáciu Box-Jenkinsová metodológia analýzy časových radov. Pri analýzach finančných časových radov často narážame na skutočnosť, že existuje korelačný vzťah medzi, po sebe nasledujúcimi, hodnotami. Potom môžeme povedať, že ide o autokorelované časové rady. Konštrukcia predpovedí týchto radov sa uskutočňuje pomocou autoregresných modelov. Pokiaľ je popis regresie založený na vzťahu susedných hodnôt, hovoríme o autoregresnom modeli 1. radu a pokiaľ je základom vzťah s posunom o dve obdobia, ide o autoregresný model 2. radu.¹⁶

Matematicky ho zapíšeme ako:

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p}$$



Obrázok 1: Predpoveď na základe autoregresie.

¹⁶ MONTGOMERY, Douglas C. – JENNINGS Cheryl L. – KULAHCI, Murat. 2008. *Introduction to Time Series Analysis and Forecasting*. 1. vydanie. New Jersey: Wiley, 2008. 472 s. ISBN 978-0471653974.

1.4.2 Proces klzavých priemerov (MA)

Model MA sa volá proces klzavých priemerov a $MA(1)$ je proces klzavých priemerov radu 1. Proces MA má tvar, v ktorom sa vyskytuje biely šum a parametre, kde stredná hodnota je nulová. Sú založené na chybnjej predikcii, sú modelmi náhodných procesov.

Matematicky ho zapíšeme ako:

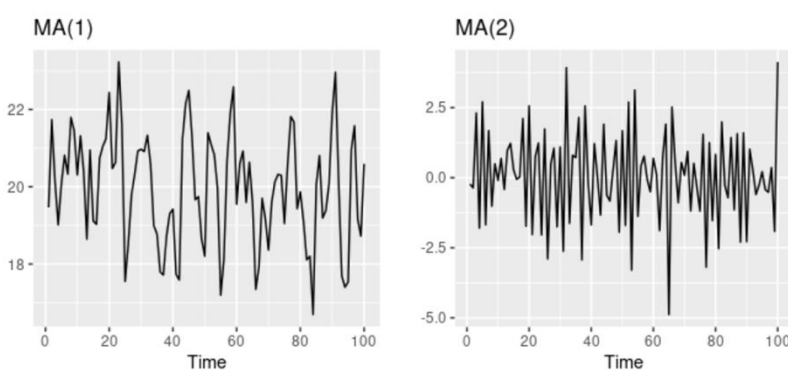
$$Y_t = \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1},$$

Kde $\theta < 1$.

Proces klzavých priemerov je jedna z možností modelovania dynamiky stacionárnych časových radov. Napríklad analýza vývoja zmien cien akcií, kde takúto zmenu ceny s nulovým priemerom a konštantným rozptylom zapíšeme ako:

$$Y_t = u_t$$

u_t sú identicky rozdelené náhodné zložky, sériovo nezkorelované. Odrážajú pôsobenie neočakávaných vplyvov na cenu akcií, napríklad informácie o finančnej situácii podniku.¹⁷



Obrázok 2: Predpoveď na základe procesu klzavých priemerov.

¹⁷ HAMILTON, Douglas James. *Times Series Analysis*. 1st ed. Princeton: Princeton University Press, 1994. 64-67 s. ISBN 978-06-9104-289-3..

1.4.3 Box-Jenkinsové modely

Box –Jenkinsové modely kladú dôraz na reziduálnu zložku modelu, pričom v nich vychádzame z predpokladu, že táto môže byť tvorená korelovanými náhodnými veličinami. Východiskovými nástrojmi týchto modelov je korelačná a regresná analýza.

1.4.3.1 Autoregressive Moving Average (ARMA)

Je kombináciou autoregresných (AR) modelov s modelmi s kľzavým priemerom (MA), ktoré sa používajú spolu na získavanie novej jednotky modelov časových radov označených ako modely ARMA ($AR + MA = \text{model ARMA}$).¹⁸

Matematicky ho zapíšeme ako:

$$x_t = z_t + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \theta_1 z_{t-1} + \dots + \theta_q z_{t-q}$$

1.4.3.2 Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA je skratka pre autoregresný integrovaný kľzavý priemer. Používa sa pri krátkodobej predikcii a to v situácií, kedy pri odhade a verifikácii dospejeme k záveru, že odhadnuté parametre sú z hľadiska ekonomických a štatistických kritérií nepoužiteľné alebo má model zlé prognostické vlastnosti.¹⁹

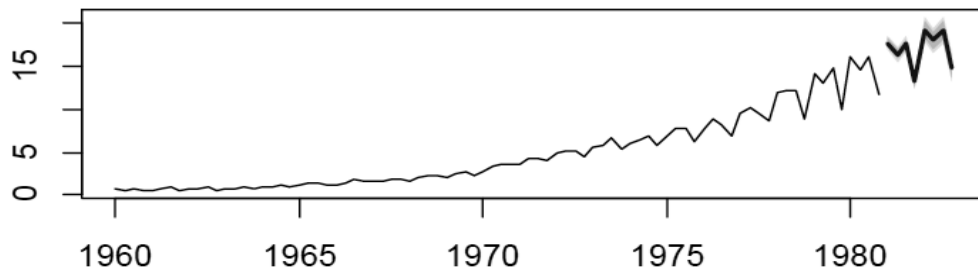
1.4.3.3 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Ak časový rad obsahuje sezónnu zložku, je najvhodnejšie použiť model SARIMA. Postup je rovnaký ako pri modeli ARIMA. Na grafe môžeme vidieť predpoveď použitím modelu SARIMA, v prvej zátvorke sú komponenty ARIMA zložky a v druhej zátvorke sú komponenty sezónnej zložky.

¹⁸ MILLS, C. Terence. *The Econometric Modelling of Financial Times Series*. 3rd ed. New York: Cambridge University Press, 2008. 10-12 s. ISBN 052171009X.

¹⁹ ANDERSEN, G. Rorben - DAVIS, A. Richard. *Handbook of Financial Times Series*. 1st ed. Berlin: Springer, 2009. 711 s. ISBN 978-3-540-71296-1.

Predpoved SARIMA (3,1,1)(0,1,0)



Obrázok 3: Predpoved' použitím modelu SARIMA.

1.4.4 Exponenciálne vyrovňovanie

Exponenciálne vyrovňovanie je metóda, ktorá patrí medzi najviac používanú metódu pre svoju výpočtovú jednoduchosť, ale najmä pre schopnosť rýchleho reagovania na posledné zmeny hodnôt, ktoré sme urobili v časovom rade. Metodika, pomocou jej vážených priemerov, prihliada na minulý vývoj sledovanej veličiny.

Exponenciálne vyrovnanie definujeme v čase t a to pomocou nasledovného vzorca:

$$S_t = \alpha y_t + \beta S_{t-1} = \beta(\alpha y_{t-1} + \beta S_{t-2}) = \alpha y_t + \alpha \beta y_{t-1} + \beta^2(\alpha y_{t-2} + \beta S_{t-3}) = \dots$$

$$= \alpha \sum_{k=0}^{N-1} \beta^k y_{t-k} + \beta^N S_0 \quad , \text{pre } k=0, 1, \dots, N-1$$

Kde:

S_t – je exponenciálny priemer v čase t

α – vyrovňavajúca konštanta z intervalu $<0,1>$; $\beta=1-\alpha$

N – dĺžka časového radu

k – vek pozorovaní

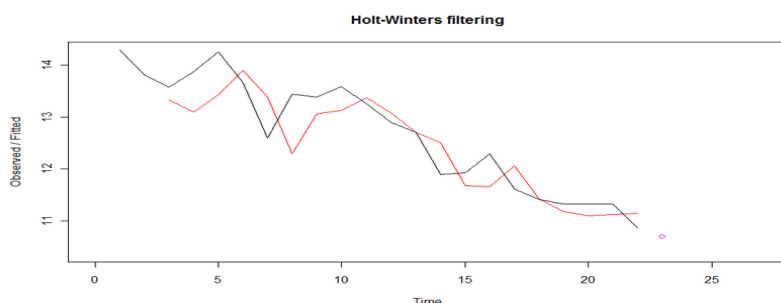
Pri použití modelu ETS sa zameriavame na vzťahy, ktoré vznikajú medzi jednotlivými zložkami. Podľa toho aký je trend, či už žiadny, lineárny alebo exponenciálny a sezónnosť, ktorá môže byť konštantná, stúpajúca alebo klesajúca v časovom rade, rozoznávame tieto druhy ETS modelu:

- a) jednoduchá metóda exponenciálneho vyrovňovania;
- b) Holtova metóda lineárnych trendov;
- c) metóda exponenciálneho trendu;
- d) Holt-Wintersova sezónna metóda.

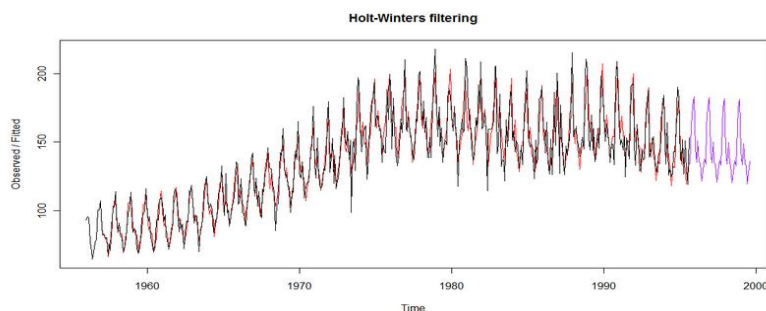
Najznámejšou funkciou exponenciálneho vyrovňovania v R je funkcia `HoltWinters`, ktorá odkáže pracovať so sezónnymi ale i nesezónnymi dátami.

Parametre funkcie:

- `x` - dáta časového radu
- `alfa`, `beta`, `gamma` – vyrovňavajúce konštanty
- sezónnosť – defaultne „additive“, pre multiplikatívny sezónny model treba zadať „multiplicative“. Tento parameter naberá účinnosť, len ak je parameter `gamma` nenulový.



Obrázok 4: Predpoveď použitím modelu exponenciálne vyrovňovanie - nesezónne dáta.



Obrázok 5: Predpoveď použitím modelu exponenciálne vyrovňovanie - sezónne dáta.

2 Cieľ a metodika práce

Analýzy, ktoré sa zaoberajú akoukoľvek kvantitatívnou analýzou finančných dát, predovšetkým štatistickým spracovaním finančných dát, medzi ktorú patrí klasická popisná štatistická analýza, štatistická identifikácia, odhad modelu a verifikácia modelu, štatistické testovanie rôznych finančných princípov a modelovanie predpovedí v rámci vytvoreného modelu, nazývame analýzy finančných časových radov.

V posledných rokoch vzniklo v oblasti analýzy ako ekonomických tak i finančných časových radov veľa nových metód a prístupov. Toto zlepšenie je spôsobené predovšetkým relatívne ľahkou dostupnosťou pomerne širokého spektra rôznych druhov kvalitných softvérov ako i rastúcou dĺžkou analyzovaných časových radov.

V našej bakalárskej práci si predstavíme, teoreticky si vymedzíme a následne aplikujeme a porovnáme základné metódy modelovania a predpovedania časových radov v rámci štatistického programového jazyka R. Metódy modelovania, na ktoré sa sústredíme sú metódy modelovania trendu za pomoci regresných funkcií, ďalej pôjde o modelovanie časových radov pomocou Autoregresných modelov ARIMA a nakoniec pomocou exponenciálneho vyrovnávania a automatických funkcií programu R.

R distribúcia obsahuje funkcie pre veľké množstvo štatistických postupov. Patria sem: lineárne a zovšeobecnené lineárne modely, nelineárne regresné modely, analýza časových radov, klasické parametrické a neparametrické testy, zhľukovanie a vyhladzovania. Je tu tiež veľká množina funkcií, ktoré poskytujú flexibilné grafické prostredie pre vytváranie rôznych typov dátových prezentácií.

V bakalárskej práci budem spracovávať, analyzovať a vyhodnocovať dva časové rady. Prvý časový rad tvorí maloobchodný predaj v USA v rokoch 1992 až 2019 a druhý časových rad tvorí HDP v Amerike v rokoch 1980 až 2016. Tieto časové rady sme si zvolili z dôvodu ľahkej dostupnosti. Pri analýze a predikcii každého časového radu bude nasledovný postup a to: urobíme si základnú charakteristiku dát pred dekompozíciou a po dekompozícii, následne si budeme modelovať dáta z časového radu, znázorníme si chybné položky z lineárnej regresie, chybnú položku ARIMA, SARIMA a ETS. Následne si vyberieme najlepšieho zástupcu z každej oblasti, ktoré si vyhodnotíme v tabuľke, ako kritérium nám poslúžia BIC a AIC kritéria. V poslednom rade si vyberieme vzorky, ktoré budeme predikovať.

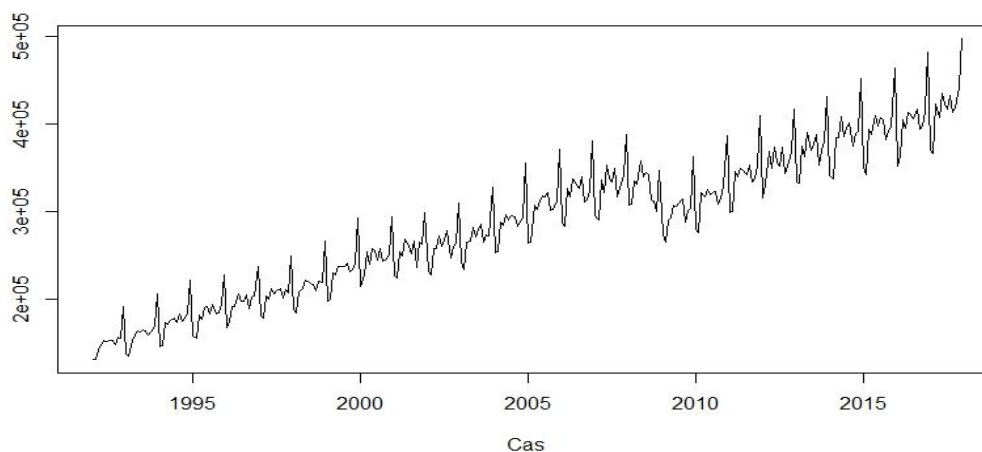
3 Výsledky práce a diskusia

V poslednej časti bakalárskej práce budeme aplikovať modely na reálne dáta finančných časových radov. Budeme spracovávať a vyhodnocovať, ktorý model má najlepšie výsledky predikcie. Na analyzovanie sme si vybrali maloobchodný predaj v Amerike v rokoch 1992 až 2019 a Americký HDP za obdobie od januára 1980 až do decembra 2016.

3.1 Analýza dát časového radu maloobchodného predaja v USA

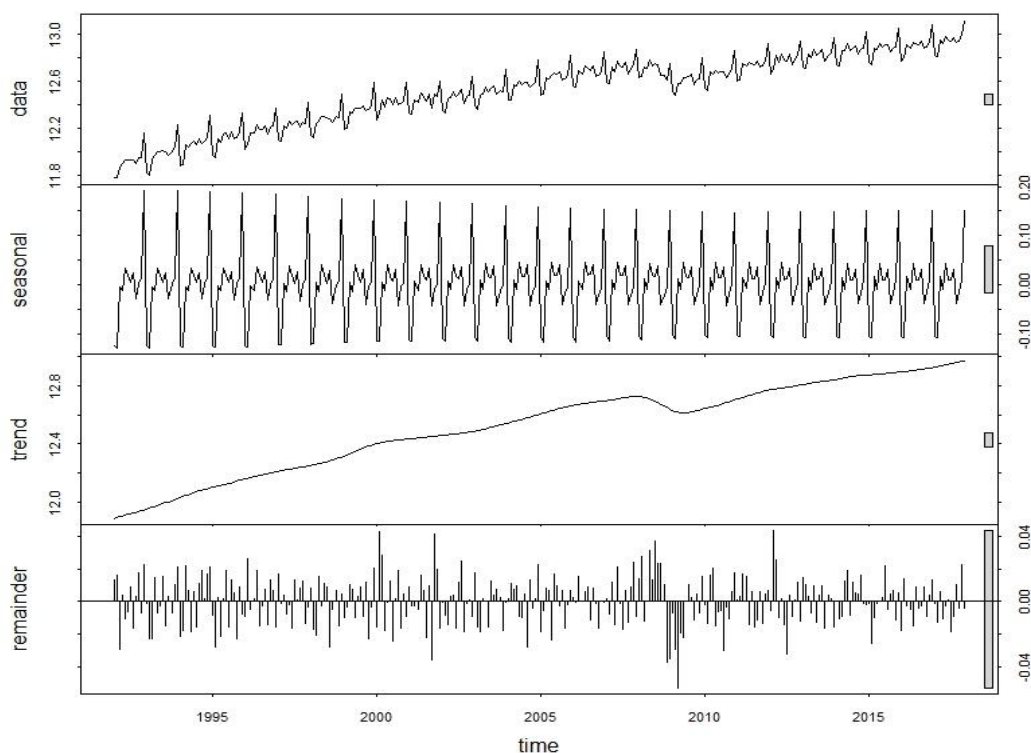
Na porovnanie, spracovanie a predikciu dát sme si vybrali maloobchodný predaj v Amerike. Maloobchod je forma obchodu, pri ktorej sa tovar predáva konečnému spotrebiteľovi. Najznámejšie maloobchody v Amerike sú napríklad Amazon, Walmart, Ebay alebo Target.

Dáta o maloobchodnom predaji sme si vybrali kvôli tomu, že sú ľahko dostupné a vykazujú krásnu sezónnosť. Spracovali sme si mesačné dáta od januára 1992 do decembra 2016. Snažíme sa predikovať dáta 3 roky dopredu, čiže o januára 2017 do decembra 2019. Toto obdobie sme si vybrali preto, lebo pandémie Covid-19 by nám predikciu rozhádzala.



Obrázok 6: Maloobchodný predaj v USA.

V prvom rade sme si znázornili graf, v ktorom vidíme jednoznačný trend, sezónnosť ako i zvyšujúcu sa volatilitu v čase. Z tohoto dôvodu je dobré, aby sme pracovali s logaritmizovanými dátami.



Obrázok 7: Rozklad časového radu.

Takto zlogaritmizované dáta si potom rozdelíme na jednotlivé komponenty. Hore vidíme našu pôvodnú, avšak teraz zlogaritmizovanú radu, pod ňou je vidieť výrazná sezónnosť. Ako tretí tu máme trend, ktorý vidíme, že je prakticky lineárny, až na krízu v roku 2008. To isté môžeme povedať aj o našej poslednej zložke, čo je to, čo zostalo. Tu môžeme opäť vidieť zvýšenú volatilitu v období krízy. Inak sa tento zvyšok javí ako náhodný.

3.1.1 Základná charakteristika dát

V nasledujúcich tabuľkách môžeme vidieť základné charakteristiky ako je maximum, minimum, stredná hodnota, medián a rozptyl pred a po dekompozícii.

Pred dekompozíciou:

Pred Dekompozíciou	Maximum	Minimum	Stredná Hodnota	1/4	Median	3/4	Rozptyl
	482 002	130 683	280 659,69	213 347,25	285 682	339 962,25	78 136,64

Tabuľka 1: Základné charakteristiky pred dekompozíciou.

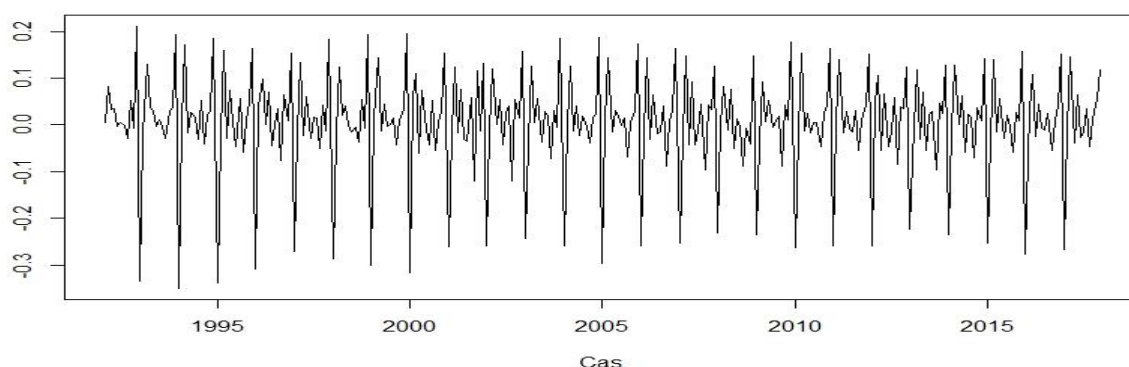
Po dekompozícii:

Log-diferencovani	Maximum	Minimum	Stredná Hodnota	1/4	Median	3/4	Rozptyl
	0,212	-0,35	0,004	-0,023	0,009	0,052	0,105

Tabuľka 2: Základné charakteristiky po dekompozícii.

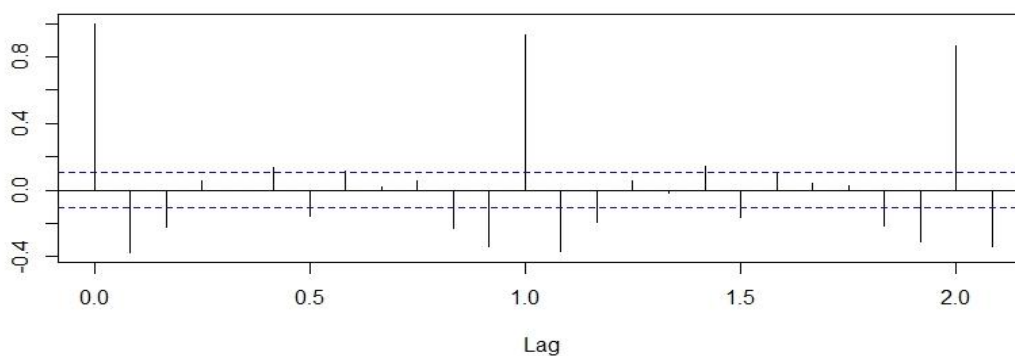
3.1.2 Modelovanie dát z časového radu

Môžeme však vidieť, že naše dáta nie sú stacionárne a tak musíme vziať zlogaritmizovaný rozdiel, aby sme túto stacionaritu dosiahli. Dickey Fuller test (-13.413 , 7 , p -hodnota 0.01) nám potom ukazuje, že stacionarita je dosiahnutá už po jednom diferencovaní a nie je potrebná ďalšia akcia (je to test, ktorý testuje, či sú dáta stacionárne, pokiaľ je p -hodnota pod $0,05$ potom sú a môžeme usudzovať stacionaritu).

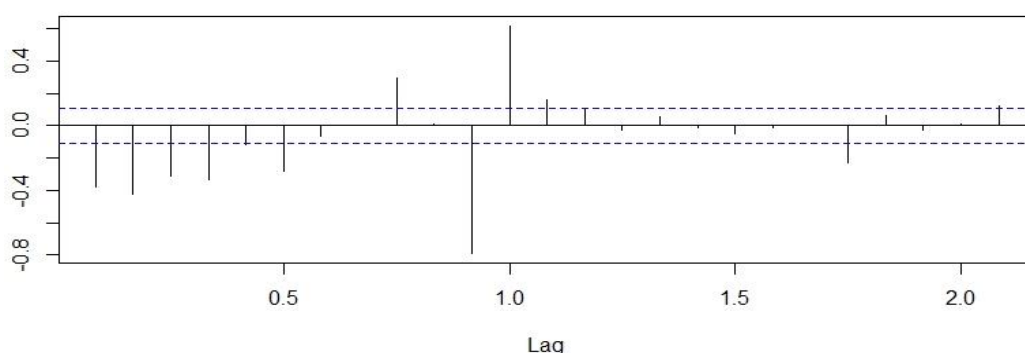


Obrázok 8: Graf časového radu po log-diferencovaní.

Ďalej máme graf takto upravených dát spoločne s ACF a PACF. Z týchto grafov môžeme vidieť, že tu nemáme len sezónnosť, ale aj nejakú tú autokoreláciu. Aby sme dosiahli presnú predikciu musíme použiť SARIMA. Z parciálnej autokorelácie vidíme, že kľzavý priemer by mal byť druhého rádu. U AR vidíme, že autokorelácia pretrváva až do 4 rádu.



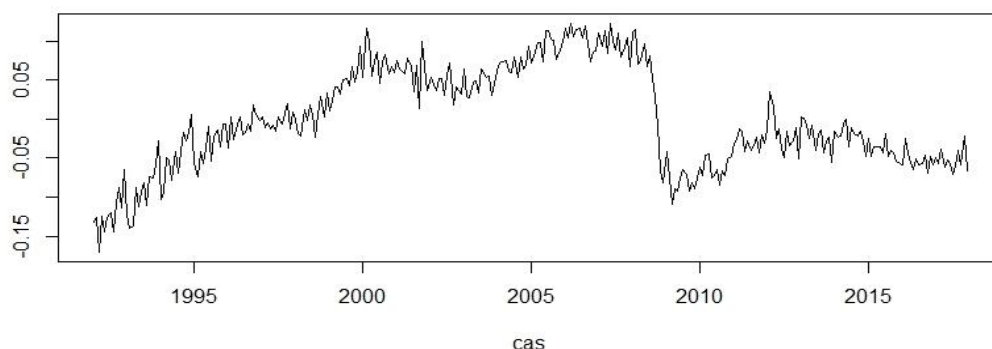
Obrázok 9: Graf ACF.



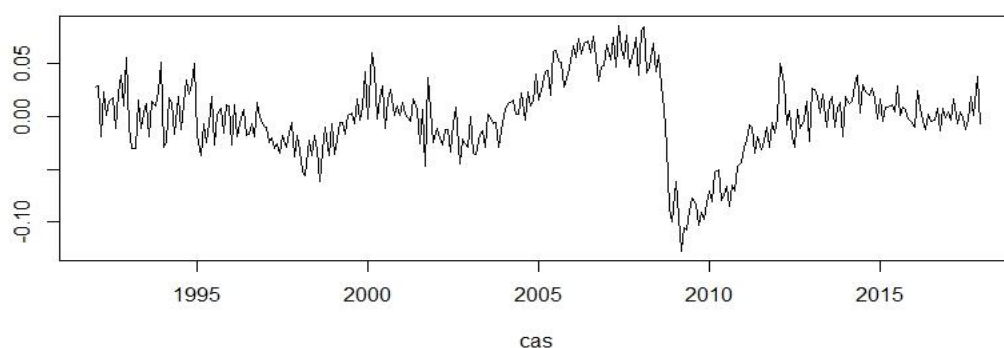
Obrázok 10: Graf PACF.

- **Lineárna regresia**

Teraz si poďme porovnať niekoľko modelov. Prvou možnosťou je lineárna regresia, kde máme trend a sezónnosť. Čiže máme jeden koeficient pre trend a 11 dummy koeficientov, pre každý mesiac mínus jeden. Musíme ich tu mať len jedenásť, inak by sme narazili na multikoreláciu a neboli by sme schopný túto regresiu vypočítať. Máme tu graf predpovede pomocou lineárnej regresie. Problém s touto metódou vidíme z obrázku 13, keď vykreslíme chybu tohto modelu. Vidíme, že nepripomína biely šum ako by mala, ale je veľmi autokorelovaná, je teda jasné že by sme mali použiť nejaký ARIMA model. Rovnaký výsledok vidíme aj z Ljung-Boxovho ($Q^*=306.13(24)$, $p=0.00$) testu, ktorý je významne signifikantný. (Test testuje, či máme autokoreláciu, pokiaľ je signifikantná, tak máme autokoreláciu, pokiaľ nie je signifikantná, tak máme biely šum). Vyskúšali sme pridať i kvadratický a kubický trend a ako môžeme vidieť v tabuľke 3 priniesol nám slabé zlepšenie i keď ako vidíme z Ljung-Boxova testu ($Q^*=247.81(24)$, $p=0.00$) i s kvadratickým a kubickým trendom máme stále veľkú autokoreláciu. Ďalšie rady trendov nám, ale nepriniesli ďalšie zlepšenie.



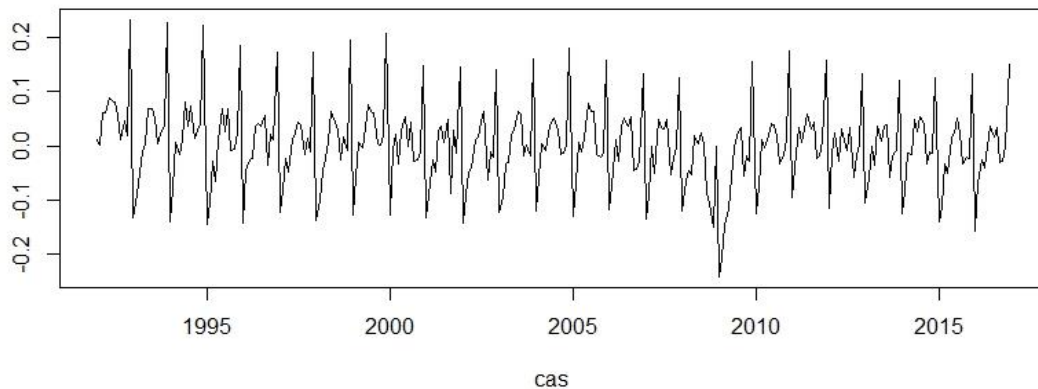
Obrázok 11: Chybná položka z lineárnej regresie.



Obrázok 12: Chybná položka lineárnej regresie s kvadratickým a kubickým trendom.

- **ARIMA**

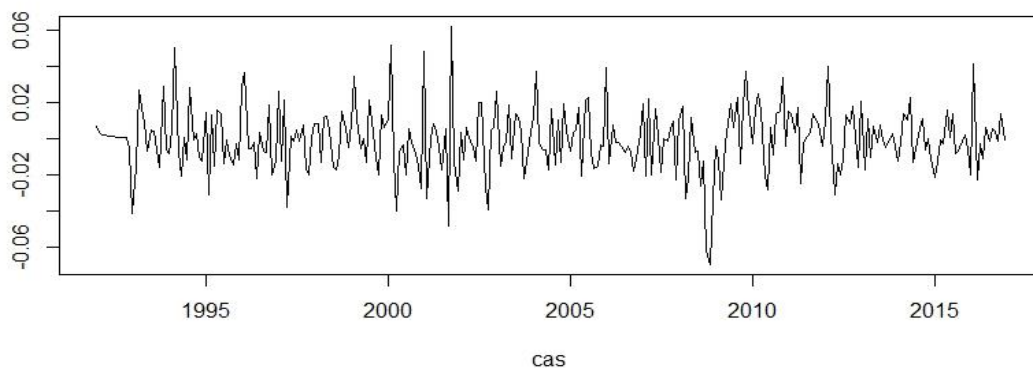
Teraz sa môžeme pozrieť na ARIMA (4,1,2) ako nám povedali ACF, PCF a Dickey Fuller test. Vieme totiž, že po jednej log-diferencii máme stacionárny rad a tento rad má 4 AR koeficienty a 2 MA koeficienty. Pokiaľ k tomu pridáme drift potom dostaneme nasledujúcu predpoveď. Vidíme však, že predpoveď má stále pomerne veľkú volatilitu, pretože nevyužívame všetky dostupné informácie a to informácie o sezónnosti. Vidíme to aj z grafu našich chýb, pretože z nich vidíme veľkú sezónnosť. Ljung-Box test nám tiež hovorí, že máme stále veľkú autokoreláciu a to práve kvôli sezónnosti ($Q^* = 487.54$ (17), $p=0.00$).



Obrázok 13: Chybná položka $ARIMA(4,1,2)$.

- **SARIMA**

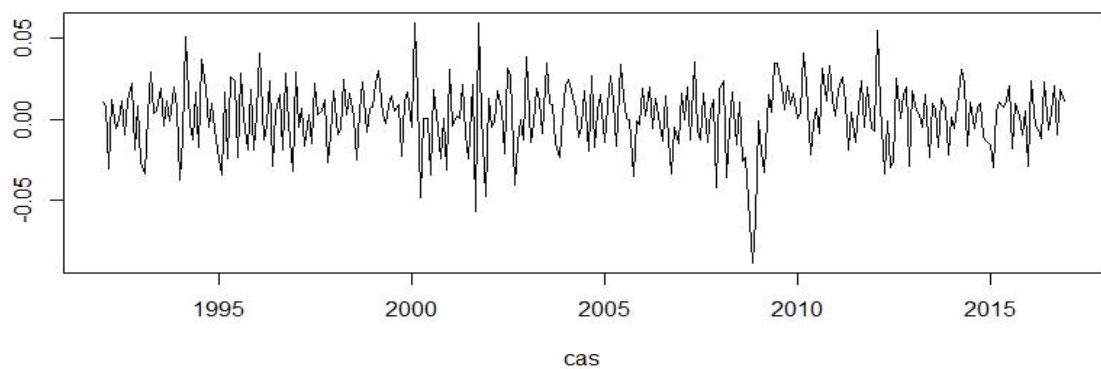
Ďalej tu máme predpoveď SARIMA. Použijeme našu pôvodnú $ARIMA$ predpoveď a necháme R vybrať sezónnu zložku. Výsledkom je model $SARIMA(4,1,2)(1,1,2)$. Vidíme, že predpoveď je podobná regresii avšak sme si oveľa istejší pri krátkodobých predpovediach než je regresia. Vidíme, že náš model pasuje na dáta výrazne lepšie, pretože naše chybné položky pripomínajú biely šum a nevidíme tu takmer žiadnu autokoreláciu. Toto nám potvrdil aj Ljung-Box test ($Q^* = 13.994(16)$, $p = 0.54$).



Obrázok 14: Chybná položka $SARIMA(4,1,2)(1,1,2)$.

- **ETS**

Tento model si ešte porovnáme s exponenciálnym vyrovňovaním. Model, ktorý vybralo R s pomocou BIC je $ETS(A,Ad,A)$. Buď máme multiplikatívnu náhodnú chybu, aditívny trend a opäť aditívnu sezónnosť. Vidíme tu opäť predpoveď a chybnú položku, ktorá tiež vyzerá podobne ako biely šum, ale je tu vidieť o trochu viac autokorelácie ako u $SARIMA$. Test Ljung-Box ($Q^* = 265.43(7)$, $p = 0.00$) nám potom potvrdil, že je tu stále autokorelácia.



Obrázok 15: Chybná položka ETS (A, Ad, A).

Predtým, ako sa pustíme do predikcií, vyberieme si najlepšieho zástupcu z každej oblasti, čiže z lineárnej regresie, ARIMA modelov a ETS. Ako kritérium nám poslúžia BIC a AIC kritéria (vždy chceme, čo najnižšiu hodnotu a to i v záporných číslach). Ako sme už povedali, z lineárnych regresíí bol najlepší model, ktorý obsahoval sezónnosť, lineárny, kubický i kvadratický trend. Rady ARIMA modelov sme ešte doplnili o model, ktorý by vybralo R, ale ako vidíme tak najlepšie si viedol model SARIMA (4,1,2) (1,1,2). Z ETS modelov tak máme zástupcu v podobe ETS (A,Ad,A), ktorý sa nám ukázal, ako model ETS s multipliktívnym trendom.

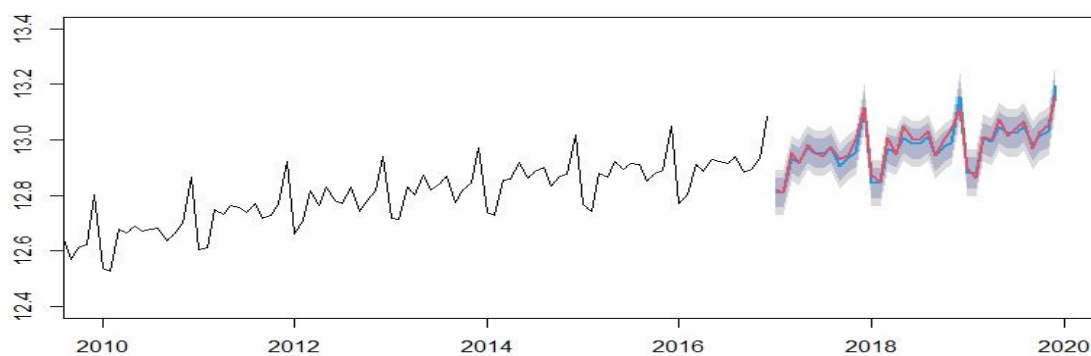
Maloobchodný predaj v Amerike	AIC	BIC
Tslm (sesonost+trend)	-789.9035	-737.501
Tslm (sesonost+trend+trend ²)	-1056.513	-1000.368
Tslm (sesonost+trend+trend²+trend³)	-1131.81	-1071.922
ARIMA(4,1,2)	-716.6925	-716.6925
SARIMA(4,1,2)(1,1,2)	-1533.887	-1496.883
SARIMA(2,1,1)(1,1,2)	-1457.347	-1483.25
ETS(M,Ad,A)	-579.8124	513.1443
ETS(A,Ad,A)	-582.0807	-515.4126

Tabuľka 3: Kritéria AIC a BIC.

3.1.3 Tréningové vzorky a ich predikcia

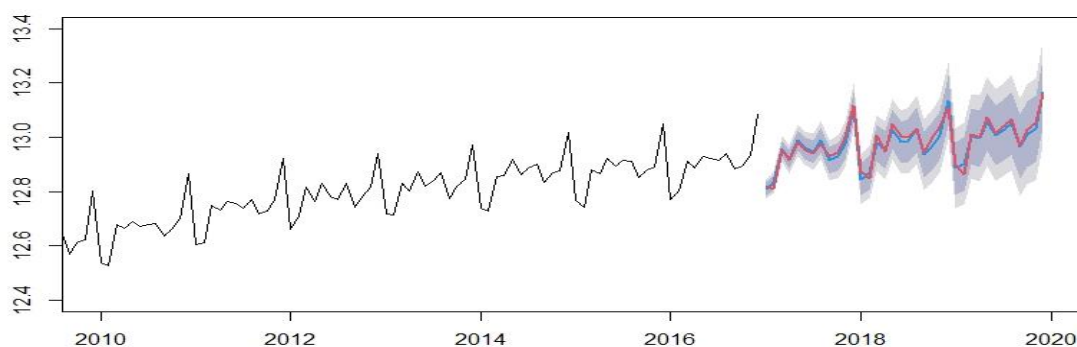
Vybrali sme si 3 modely a pôjdeme sa pozrieť ako si viedli v našej predikcii.

- *Lineárna regresia*



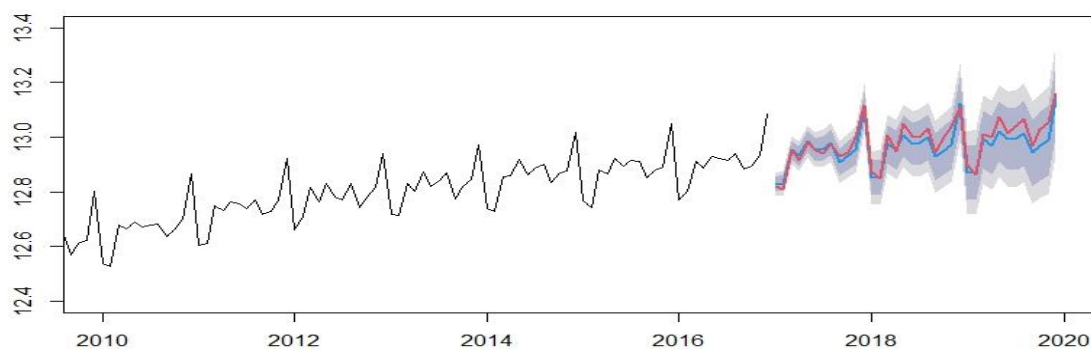
Obrázok 16: Predpoveď TSLM (sezónnosť + trend + trend2 + trend3).

- *ARIMA*



Obrázok 17: Predpoveď ARIMA (4,1,2) (1,1,2).

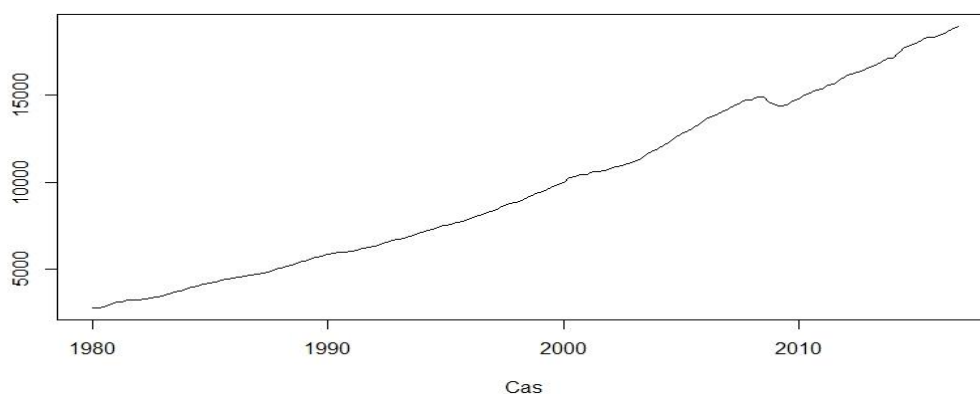
- *ETS*



Obrázok 18: Predpoveď ETS (A, Ad, A).

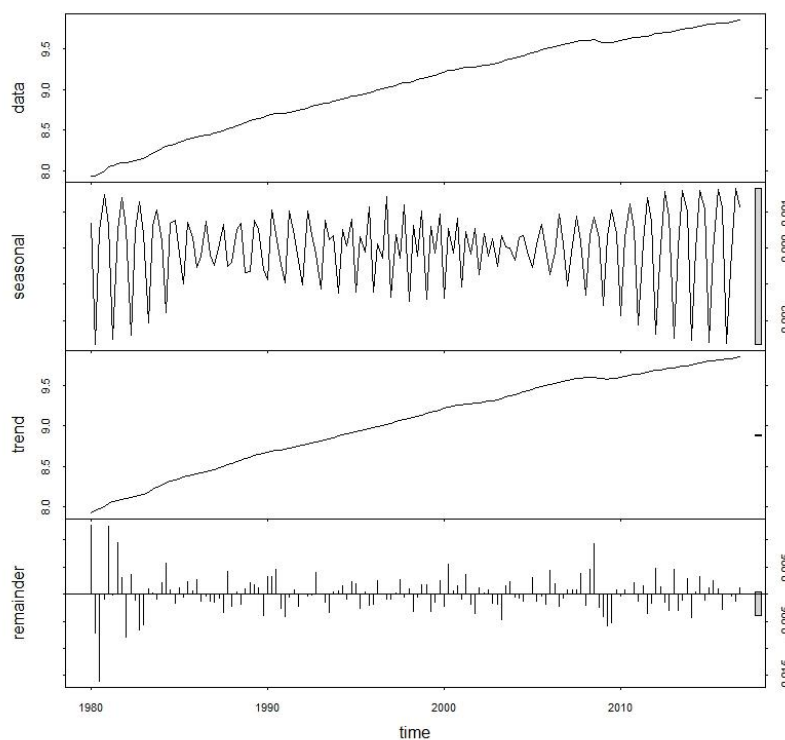
3.2 Analýza dát časového radu HDP v Spojených štátoch amerických

Ako druhú vzorku sme si vybrali časový rad hodnôt HDP v Spojených štátoch amerických. Spojené štáty americké majú jednu z najsilnejších ekonomík na svete a taktiež najpoužívanejšiu menu pri medzinárodných platbách. V našej práci sme si spracovali mesačné dáta od januára 1980 do decembra 2016. Tieto dáta sme si vybrali z dôvodu, že sú tiež ako maloobchodný predaj dobre a ľahko dostupné, sú dôveryhodné a nemajú jasnú sezónnosť, takže nám poskytnú krásny prototyp k našim prvým pozorovaným dátam.



Obrázok 19: HDP (USA).

Na obrázku vidíme jednoznačný exponenciálny trend.



Obrázok 20: Rozklad časového radu.

3.2.1 Základná charakteristika dát

V nasledujúcich tabuľkách môžeme vidieť základné charakteristiky ako je maximum, minimum, stredná hodnota, medián a rozptyl pred a po dekompozícii.

Pred dekompozíciou:

Pred dekompozíciou	Maximum	Minimum	Stredná hodnota	1/4	Medián	3/4	Rozptyl
	18 968,04	2789,84	9776,05	5587,16	9045,4	14409,29	4827,44

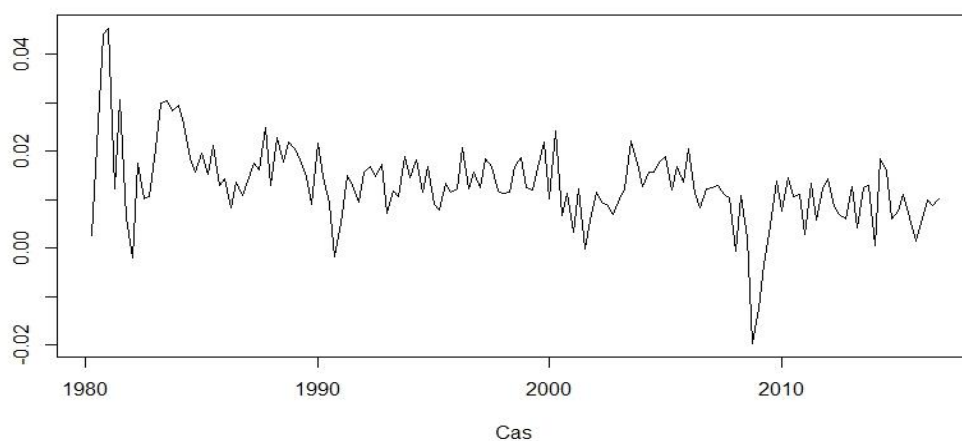
Tabuľka 4: Základné charakteristiky pred dekompozíciou.

Po dekompozícii:

	Maximum	Minimum	Stredná hodnota	1/4	Medián	3/4	Rozptyl
Prvé Log-diferencovanie	0,045	-0,02	0,013	0,009	0,012	0,017	0,008
Druhé Log-diferencovanie	0,023	-0,033	0	-0,005	0	0,005	0,008

Tabuľka 5: Základné charakteristiky po dekompozícii.

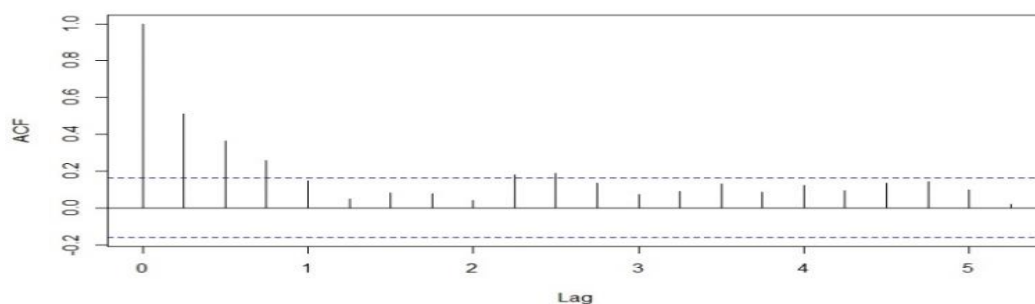
3.2.2 Modelovanie dát z časového radu



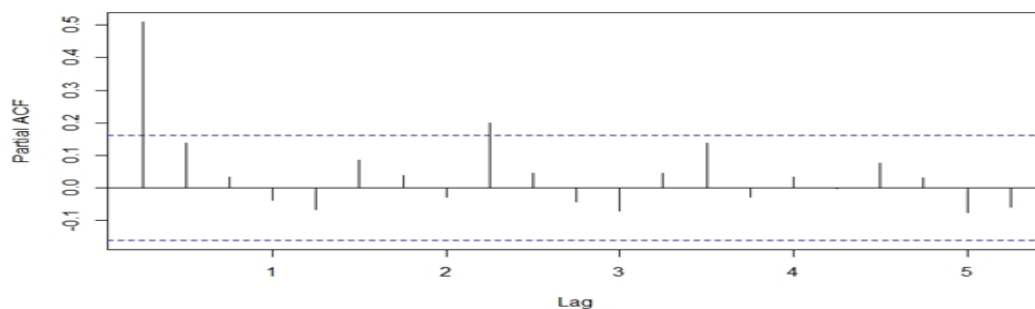
Obrázok 21: Graf časového radu po log-diferencovaní.

Tento graf nám veľa nepovie. Poďme si vziať teda prvú diferenciu a pozrieť sa, či je rada stacionárna. Dikey Fuller test (-5.00629 (5) $p=0.01$) nám ukazuje, že stacionarita je dosiahnutá už po prvom diferenciovaní a nie je potrebná ďalšia akcia.

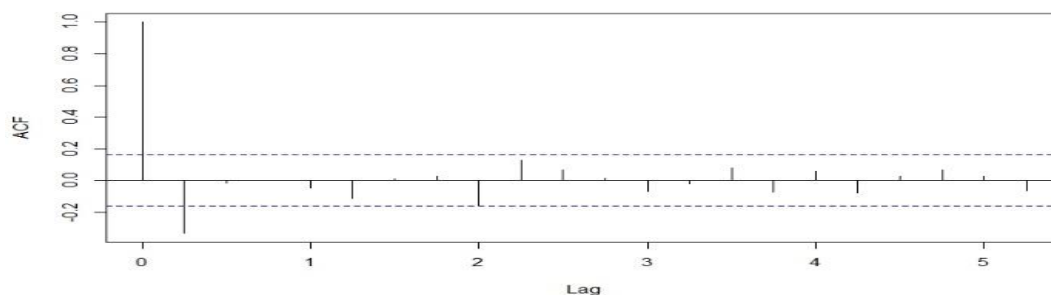
Takto potom vyzerajú naše stacionárne dáta. Vidíme silnú autokoreláciu v našich dátach a nepravidelnosť. Nemáme tu žiadnu jasnú sezónnosť ako v predchádzajúcej analýze. Rovnako vidíme, že hoci test stacionarity nám signalizuje stacionaritu v dátach, vidíme nepatrný negatívny trend. Bude preto zaujímavé porovnať model s prvou a druhou diferenciou. Poďme však teraz k prvej diferencii.



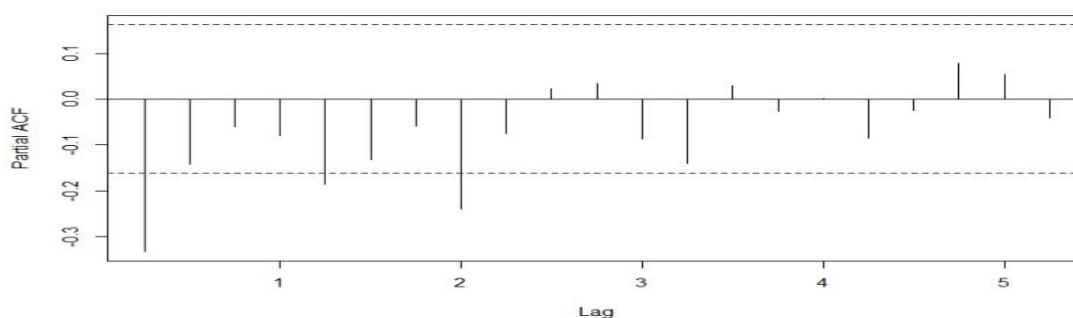
Obrázok 22: ACF po prvom log-diferencovaní.



Obrázok 23: PACF po prvom log-diferencovaní.



Obrázok 24: ACF po druhom log-diferencovaní.

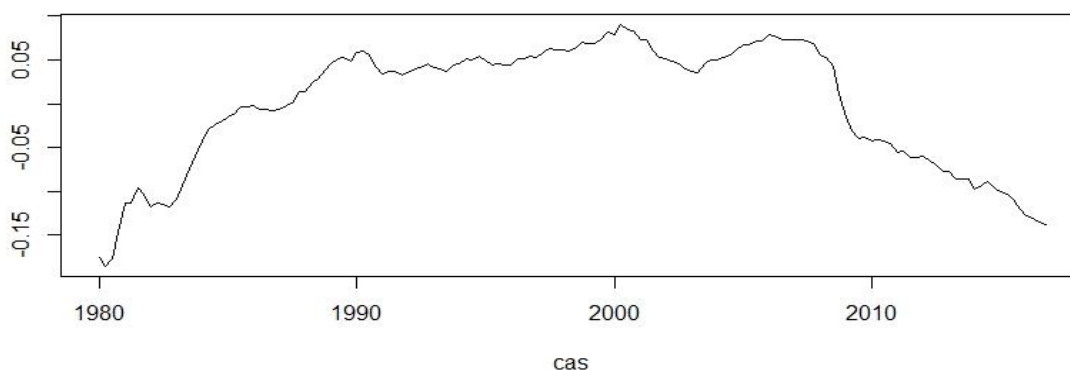


Obrázok 25: PACF po druhom log-diferencovaní.

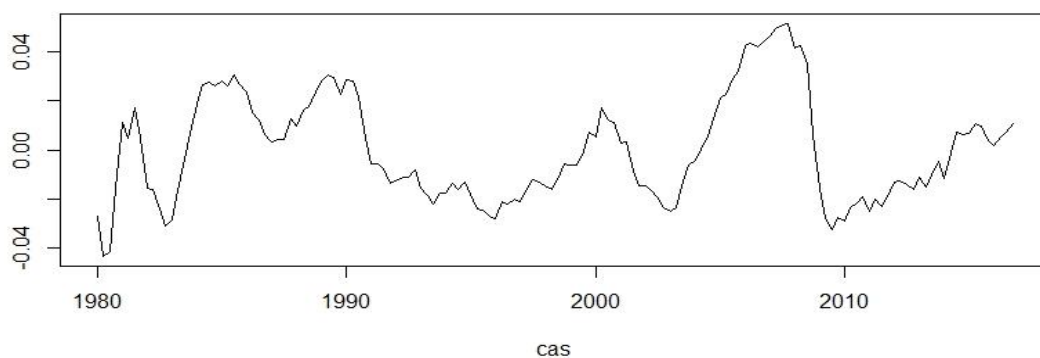
Z ACF a PACF grafov vidíme, že tu máme štyri rady MA a ani jednu radu AR. Môžeme tu však vidieť nejaké ľahké autokorelácie aj v neskorších periodách. Poďme sa pozrieť ako vyzerajú grafy pre dvojito diferencovanú radu. Vidíme, že tu máme jasný jeden rad MA. S AR procesmi to nie je také jednoznačné, nakoľko ako vidíme z grafu PACF máme tu stále pomerne silnú autokoreláciu aj v ďalších periodách. Nechali sme teda program R, aby vybral najviac vhodný počet autokorelácií a ten sa ukázal byť 1.

- **Lineárna regresia**

Poďme si opäť porovnať ako jednotlivé modely odpovedajú našim dátam. Z dôvodu, že z našich dát nie je vidieť žiadna sezónnosť ako prvú sme urobili lineárnu regresiu iba s trendom. Z Ljung-Box testu ($Q^*=140.49$ (8), $p=0.00$) i grafu chybné položky vidíme, že chybná položka má stále silnú autokoreláciu a čo je najviac, vidíme tu aj jasný kvadratický trend. Keď použijeme regresiu i s kvadratickým trendom, tak z Ljung-Box testu $Q^*=133,88$ (8), $p=0.00$) i grafu vyplýva, že tu máme stále silnú autokoreláciu v našej chybovej položke. Pridaním ďalších radov trendu už ale nie sme schopní dosiahnuť pokroku a ako vidíme z tabuľky 6, tak ani pridaním sezónnosti.



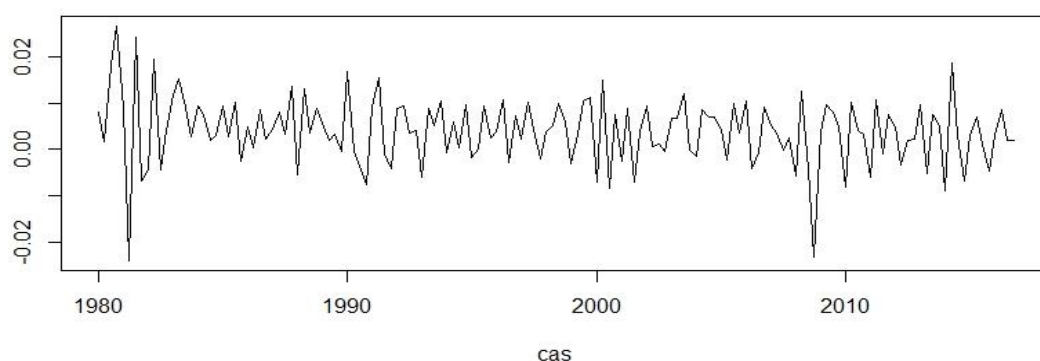
Obrázok 26: Chybná položka z lineárnej regresie s lineárnym trendom.



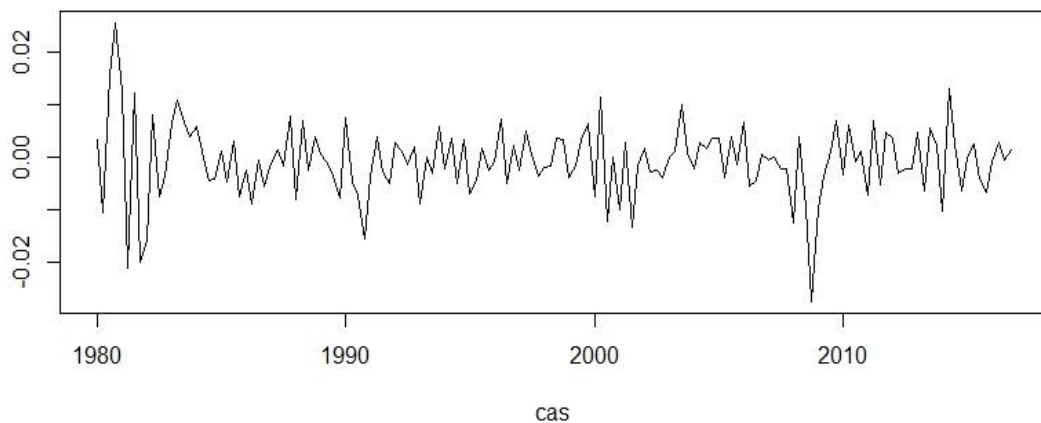
Obrázok 27: Chybná položka z lineárnej regresie s kvadratickým trendom.

- **ARIMA**

Keďže dáta stále vykazovali autokoreláciu, poďme sa pozrieť na ARIMA modely. Najprv poďme bližšie preskúmať model, ktorý sme vybrali prvým diferencovaním a to model ARIMA (0,1,4) so štyrmi MA procesmi. Z grafov je zrejmé, že naša chybná položka je veľmi podobná bielemu šumu ako u predchádzajúcich modeloch. Vidíme tu len miernu autokoreláciu a odchylky sa v našom modeli zdajú byť náhodné. Ljung-Box test ($Q^*=18.366$, (4), $p=0.00$) nám však stále zamieta nulovú hypotézu žiadnej autokorelácie v našich dátach. Pre model ARIMA (1,2,1) sa naša chybná položka chová veľmi podobne i keď sa pozrieme na Ljung-Box test ($Q^*=112.799$, (4), $p=0.06$), vidíme, že miera autokorelácie je o niečo málo nižšia ako pre predchádzajúci model. (p-hodnota je 0.06 čiže nad hodnotou 0.05 čo značí, že už nezamietame nulovú hypotézu a môžeme povedať, že naša chybná položka už nemá autokoreláciu, ale ako vidíme, tak len veľmi tesne, preto by sme mali byť s interpretáciou veľmi opatrní).



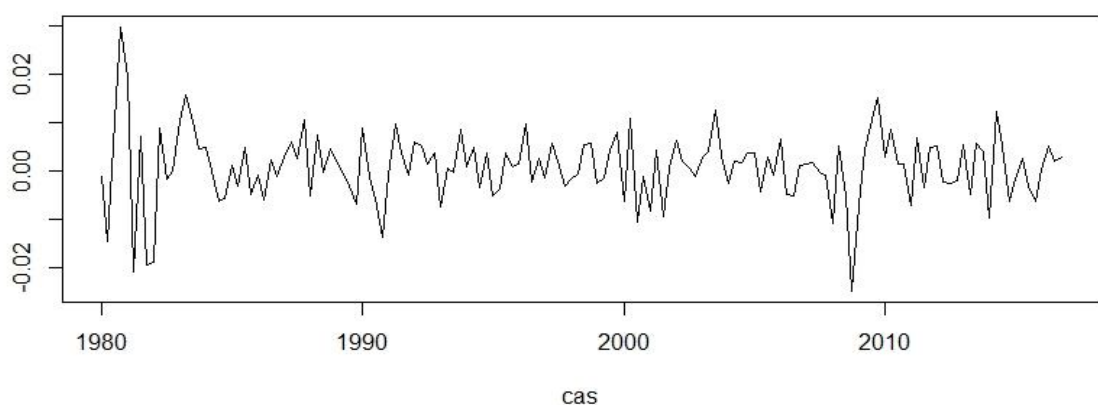
Obrázok 28: Chybná položka ARIMA (0,1,4).



Obrázok 29: Chybná položka $ARIMA(1,2,1)$.

- **ETS**

Pre ETS model sme vybrali modely ETS (A, Ad, N), čiže rovnaký model ako pri maloobchodnom predaji, ale bez sezónnosti. Vidíme aj tu, že naša chybná položka pripomína biely šum a že aj tu máme iba malú autokoreláciu v tej chybnnej položke. Aj keď, ako vidíme z Ljung-Box testu ($Q^*=15.309$, (3), $p=0.00$), tak stále nejakú máme. Rovnako ako aj u prechádzajúcich ARIMA modelov vidíme ľahkú heteroskedasticitu (to znamená, že tá chybná odchylka nemá po celý čas rovnakú variáciu, ale v niektorých obdobiach, hlavne v dobe finančnej krízy, je vyššia ako v iných), toto je však normálne pri finančných dátach a nemá zmysel sa tým extra zaoberať.



Obrázok 30: Chybná položka (A, Ad, N).

V nasledujúcej tabuľke vidíme, ako si jednotlivé modely vedli podľa AIC a BIC kritérií. Vidíme, že pre lineárnu regresiu je najlepší model s kvadratickým trendom, pre ARIMA modely si nakoniec najlepší viedol model s dvojitém diferencovaním, čiže ARIMA (1,2,1) a za ETS bol vybraný model ETS (A, Ad, N).

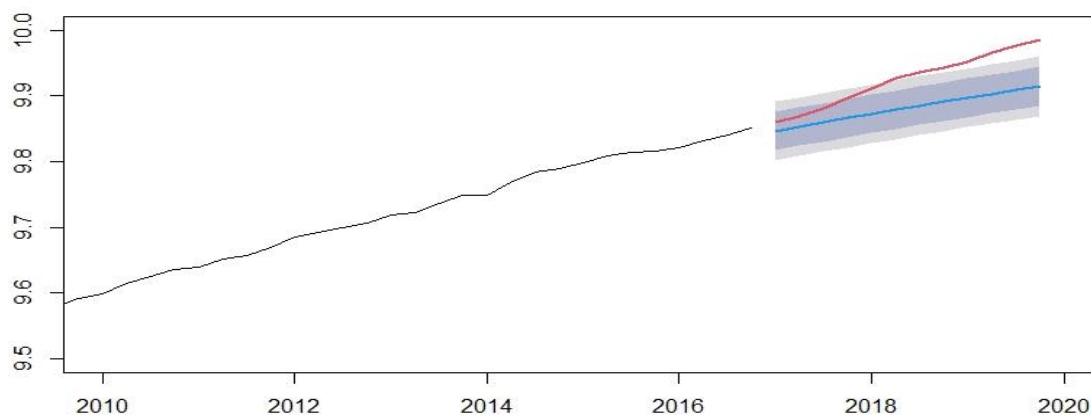
HDP v Spojených štátoch amerických	AIC	BIC
Tslm (trend)	- 357.546	-348.5544
Tslm (trend+trend²)	-708.585	-696.5961
Tslm (trend+trend ² +trend ³)	-707.0171	-692.031
Tslm (sesonost+trend+trend ²)	-702.7339	-681.7534
ARIMA(0,1,4)	-980.7814	-965.8292
ARIMA(1,2,1)	-1030.132	-1021.182
ETS(A,Ad,N)	-707.8283	-689.8450

Tabuľka 6: Kritéria AIC a BIC.

3.2.3 Tréningové vzorky a ich predikcia

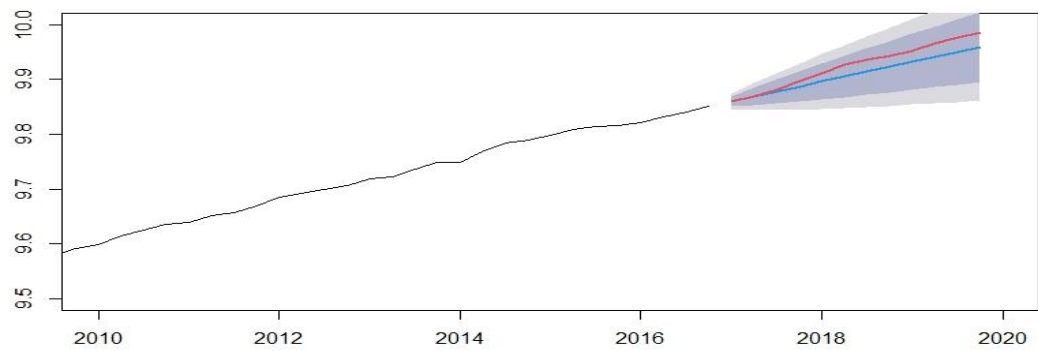
Nižšie máme zobrazené jednotlivé predikcie. Môžeme vidieť, že všetky predikcie podhodnotili HDP medzi rokmi 2017 – 2019. Vidíme tiež, že náš ARIMA (1,2,1) model bol najbližšie k skutočnej hodnote zatiaľ, čo lineárna regresia neobstála.

- *Lineárna regresia*



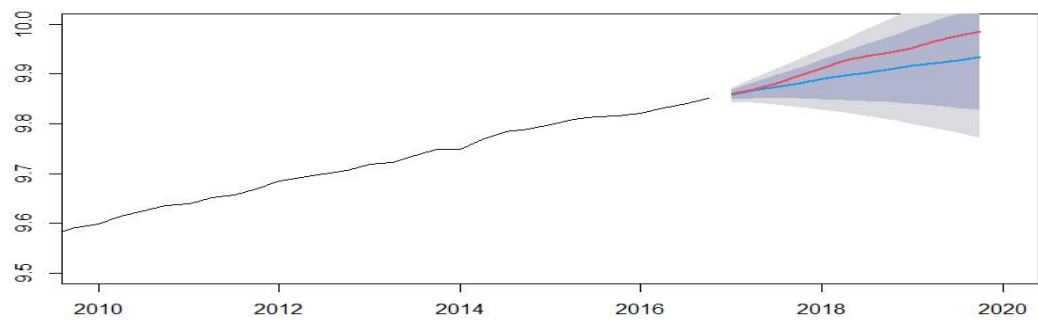
Obrázok 31: Predpoveď TSLM (trend + trend²).

- *ARIMA*



Obrázok 32: *Predpoved' ARIMA (1,2,1).*

- *ETS*



Obrázok 33: *Predpoved' ETS (A, Ad, N).*

3.3 Hodnotenie výsledkov

Pokiaľ porovnáme, ako veľmi sa líšila predikcia od reality pre jednotlivé modely vidíme, že v oboch prípadoch si najlepší viedol model zo skupiny ARIMA modelov, ako pre maloobchodný predaj tak aj pre HDP. Práve tieto modely mali najmenší RMSE. K nášmu prekvapeniu nám prepadli modely ETS, kedy pre maloobchodný predaj sa ukázali byť najmenej presné dokonca ako i obyčajný rozklad na jednotlivé zložky pomocou lineárnej regresie. (tu by mohlo byť zaujímavé zamyslieť sa nad rozdielmi medzi ARIMA a ETS a prečo v našich dátach funguje lepšie ARIMA).

Maloobchodný predaj v Amerike	ME	RMSE	MAE	MPE	MAPE	MASE
Linearni Regresse	0.0093	0.0220	0.0175	0.0719	0.1346	0.3393
SARIMA(4,1,2)(1,1,2)	0.0058	0.0175	0.0143	0.0443	0.1101	0.2768
ETS(A,Ad,A)	0.0196	0.0315	0.0255	0.1506	0.1963	0.4948

HDP	ME	RMSE	MAE	MPE	MAPE	MASE
Linearni Regresse	0.0441	0.0480	0.0441	0.4432	0.4432	0.8247
ARIMA(1,2,1)	0.0159	0.0186	0.0159	0.1595	0.1597	0.2973
ETS(A,Ad,N)	0.0272	0.0320	0.0272	0.2735	0.2735	0.5093

Tabuľka 7: Vyhodnotenie výsledkov.

Záver

Bakalárska práca sa zoberá časovými radmi, ktoré už sami o seba nesú množstvo informácií o správaní a vývoji pozorovaného javu v danom období.

Cieľom tejto bakalárskej práce bolo modelovanie vybraných časových radov, a to maloobchodného predaja v USA a Amerického HDP v určitých rokoch. Pomocou vhodného modelu bolo možné previesť aj ich predikciu. Analýza bola urobená za pomoci spracovanej teórie, podľa ktorej sme aplikovali hlavné postupy a modely na nami vybrané časové rady. Vďaka týmto postupom a metódam boli overené a otestované vybrané dáta a následne bola určená predikcia týchto dvoch časových radov.

Bakalársku prácu sme rozdelili do troch kapitol. V prvej kapitole sme sa oboznámili s časovými radmi, definovali sme si ich. Dozvedeli sme sa, aké poznáme rozdelenie časových radov, oboznámili sme sa s jednotlivými zložkami, ktoré tvoria časový rad. Pozreli sme sa na analýzu časových radov, charakterizovali sme si finančné časové rady a povedali sme si niečo aj o ich vlastnostiach. V ďalšej časti sme sa venovali predikcii a postupu pri predikovaní časových radov. Dozvedeli sme sa aké metódy sa používajú pri predikcii časových radov a tieto metódy si bližšie popísali.

Druhá kapitola bakalárskej práce bola venovaná cieľom a metodike bakalárskej práce. Oboznámili sme sa v nej s akým programom budeme pri predikcii časových radov pracovať, a popísali sme si aké modely budeme v našej práci aplikovať pri predikcii časových radov.

Tretia kapitola bakalárskej práce bola zameraná a venovaná samotnému porovnaniu dvoch vybraných časových radov, ktoré sme si charakterizovali, modelovali a predikovali a v poslednom rade porovnali a vyhodnotili.

Pokiaľ porovnáme v bakalárskej práci, ako veľmi sa líšila predikcia od reality pre jednotlivé modely vidíme, že v oboch prípadoch si najlepšie viedol model zo skupiny ARIMA modelov, ako pre maloobchodný predaj tak aj pre HDP.

Zoznam použitej literatúry:

- [1] ANDERSEN G. Rorben - DAVIS, A. Richard. *Handbook of Financial Times Series*. 1st ed. Berlin: Springer, 2009. 1050 s. ISBN 978-3-540-71296-1.
- [2] ANDERSON, David R. et al. *Statistics for Bussiness and Economics China*: RR Donnelley. 2015. 516 s. ISBN 978-1-4080-72223-3.
- [3] ARLT, Jozef - ARLTOVÁ Markéta. *Ekonomické časové řady*. vyd. 1. Praha: Profesional Publishing a.s., 2009. 290 s. ISBN 978-80-8694-685-6.
- [4] ARLT, Josef - ARLTOVÁ Markéta - RUBLÍKOVÁ, Eva. *Analýza ekonomických časových řad a příklady*. 1. vydanie. Praha: Vysoká škola ekonomická Praha, 2002. 148 s. ISBN 80-245-0307-7.
- [5] ARLT, Josef - ARLTOVÁ, Markéta. *Finančný časové řady : vlastnosti, metody modelování, příklady a aplikace*. Praha : Grada Publishing a.s., 2003. 220 s. ISBN 80-247-0330-0.
- [6] CIPRA, Tomáš. *Finanční a pojistné vzorce*. vyd. 1 Praha: Grada Publishing a.s., 2006. 374 s. ISBN 80-247-1633-X.
- [7] CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii: celostátní vysokoškolská učebnice pro stud. Matematicko-fyz. fakulty studijních odb. 11 Fyzik. matematické vědy*. Praha: Státní nakladatelství technické literatury, 1986. 246 s.
- [8] HAMILTON, Douglas James. *Times Series Analysis*. 1st ed. Princeton: Princeton Review - University Press, 1994. 820 s. ISBN 978-06-9104-289-3.
- [9] HINDLS, Richard - HRONOVÁ, Stanislava - SEGER Jan. *Statistika pro ekonomy*. vyd. 7. Praha: Profesional Publishing, 2002. 415 s. ISBN 80-864-1930-4.
- [10] MILLS, C., Terence. *The Ekonometric Modelling of Financial Times Series*. 3rd ed. New York: Cambridge University Press, 2008. 472 s. ISBN 052171009X.
- [11] MONTGOMERY, Douglas C. – JENNINGS Cheryl L. – KULAHCI, Murat. *Introduction to Time Series Analysis and Forecasting*. 1. vydanie. New Jersey: Wiley, 2008. 643 s. ISBN 978-0471653974.
- [12] SIVÁK, Rudolf a kol. *Financie*. Bratislava: Wolters Kluwer, 2019. 440 s. ISBN 978-80-7598-533-0.

[13] SODOMOVÁ, Eva a kol. *Štatistika pre bakalárov*. 2. vyd. Bratislava : EKONÓM, 2013. 246 s. ISBN 978-80-225-3614-1.

Elektronické dokumenty:

[14] HYNDMAN, J. Rob – ATHANASOPOULOS, George. *Forecasting: Principles and Practice*. 2. vydanie. [elektronický zdroj]. 2. vydanie Melbourne : OTexts. 2018. ISBN 978-0987507112. [cit. 23.10.2021]. Dostupné na: <https://otexts.com/fpp2>.

Zoznam ilustrácií

Obrázok 1: Predpoveď na základe autoregresie.....	18
Obrázok 2: Predpoveď na základe procesu kĺzavých priemerov.....	19
Obrázok 3: Predpoveď použitím modelu SARIMA.....	21
Obrázok 4: Predpoveď použitím modelu exponenciálne vyrovňavanie - nesezónne dáta..	22
Obrázok 5: Predpoveď použitím modelu exponenciálne vyrovňavanie - sezónne dáta.	22
Obrázok 6: Maloobchodný predaj v USA.	24
Obrázok 7: Rozklad časového radu.....	25
Obrázok 8: Graf časového radu po log-diferencovaní.	26
Obrázok 9: Graf ACF.	27
Obrázok 10: Graf PACF.	27
Obrázok 11: Chybná položka z lineárnej regresie.	28
Obrázok 12: Chybná položka lineárnej regresie s kvadratickým a kubickým trendom.	28
Obrázok 13: Chybná položka ARIMA (4,1,2).....	29
Obrázok 14: Chybná položka SARIMA (4,1,2) (1,1,2).	29
Obrázok 15: Chybná položka ETS (A, Ad, A).	30
Obrázok 16: Predpoveď TSLM (sezónnosť + trend + trend2 + trend3).	31
Obrázok 17: Predpoveď ARIMA (4,1,2) (1,1,2).	31
Obrázok 18: Predpoveď ETS (A, Ad, A).	31
Obrázok 19: HDP (USA).....	32
Obrázok 20: Rozklad časového radu.....	32
Obrázok 21: Graf časového radu po log-diferencovaní.	33
Obrázok 22: ACF po prvom log-diferencovaní.	34
Obrázok 23: PACF po prvom log-diferencovaní.	34
Obrázok 24: ACF po druhom log-diferencovaní.	34
Obrázok 25: PACF po druhom log-diferencovaní.	35
Obrázok 26: Chybná položka z lineárnej regresie s lineárnym trendom.	35
Obrázok 27: Chybná položka z lineárnej regresie s kvadratickým trendom.	36
Obrázok 28: Chybná položka ARIMA (0,1,4).....	36
Obrázok 29: Chybná položka ARIMA (1,2,1).....	37
Obrázok 30: Chybná položka (A, Ad, N).	37
Obrázok 31: Predpoveď TSLM (trend + trend2).	38
Obrázok 32: Predpoveď ARIMA (1,2,1).....	39
Obrázok 33: Predpoveď ETS (A, Ad, N).	39

Zoznam tabuliek

Tabuľka 1: Základné charakteristiky pred dekompozíciou.	25
Tabuľka 2: Základné charakteristiky po dekompozícii.	26
Tabuľka 3: Kritéria AIC a BIC.	30
Tabuľka 4: Základné charakteristiky pred dekompozíciou.	33
Tabuľka 5: Základné charakteristiky po dekompozícii.	33
Tabuľka 6: Kritéria AIC a BIC.	38
Tabuľka 7: Vyhodnotenie výsledkov.	40

Prílohy

Príloha č. 1: Zdrojový kód na predikciu časového radu maloobchodného predaja v USA.

```
library(forecast)
library(lmtest)
library(car)
library(tseries)

Retail_Sale<-ts(data$RSXFSN, start=c(1992,1), end = c(2022,3),
frequency = 12)
Retail_Sales<-window(Retail_Sale, end = c(2016, 12))
Retail_Sales_train<-window(Retail_Sale, start= c(2017,1),end =
c(2019, 12))

plot(Retail_Sales, ylab=NA, xlab="Cas")
acf(Retail_Sales)
pacf(Retail_Sales)
n<-stl(log(Retail_Sales),s.window = 12)
plot(n)

RS<-diff(log(Retail_Sales))
RS2<-diff(RS)
adf.test(RS)
plot(RS, ylab=NA, xlab="Cas")
acf(RS, main=NA)
pacf(RS, main=NA)

m4<-Arima(log(Retail_Sales),c(4,1,2),c(1,1,2))
summary(m4)
checkresiduals(m4)
plot(m4$residuals, ylab=NA, xlab="cas")
m3<-Arima(log(Retail_Sales),c(4,1,2),include.drift = T )
summary(m3)
plot(m3$residuals, ylab=NA, xlab="cas")
checkresiduals(m3)
```

```

m2<-auto.arima(log(Retail_Sales))
summary(m2)
plot(m2$residuals, ylab=NA, xlab="cas")
checkresiduals(m2)

m5<-forecast(ets(log(Retail_Sales),model="AAA",damped=T), h=36)
summary(m5)
checkresiduals(m5)
plot(m5$residuals, ylab=NA, xlab="cas")

m1<-tslm(log(Retail_Sales)~trend+I(trend^2)+I(trend^3)+ season)
summary(m1)
checkresiduals(m1)
plot(m1$residuals, ylab=NA, xlab="cas")

AIC(m1)
BIC(m1)

a<-accuracy(forecast(m5,h=36),log(Retail_Sales_train))
tab_df(as.data.frame(a),digits=4)

plot(forecast(m5,h=36),
      xlim=c(2010,2020),ylim=c(12.4,13.4),main="")
lines(log(Retail_Sales_train), col=2, lwd=2)

```

Príloha č. 2: Zdrojový kód na predikciu časového radu Hrubého domáceho produktu v Spojených štátoch Amerických.

```

library(forecast)
library(lmtest)
library(car)
library(tseries)

GDP<-ts(data2$GDP, start=c(1947,1), end = c(2022,1), frequency =
4)
GDP_train<-window(GDP, start= c(1980,1),end = c(2016, 4))

```

```

GDP_test<-window(GDP, start= c(2017,1),end = c(2019, 4))

plot(GDP_train, ylab=NA, xlab="Cas")
acf(Retail_Sales)
pacf(Retail_Sales)
n<-stl(log(GDP_train),s.window = 12)
plot(n)

RS<-diff(log(GDP_train))
RS2<-diff(RS)
adf.test(RS)
plot(RS, ylab=NA, xlab="Cas")
acf(RS2, main=NA)
pacf(RS2, main=NA)

m4<-Arima(log(Retail_Sales),c(4,1,2),c(1,1,2))
summary(m4)
checkresiduals(m4)
plot(m4$residuals, ylab=NA, xlab="cas")
m3<-Arima(log(GDP_train),c(0,1,4))
summary(m3)
plot(m3$residuals, ylab=NA, xlab="cas")
checkresiduals(m3)

m2<-Arima(log(GDP_train),c(0,2,1))
summary(m2)
plot(m2$residuals, ylab=NA, xlab="cas")
checkresiduals(m2)

m5<-forecast(log(GDP_train), h=12)
summary(m5)
checkresiduals(m5)
plot(m5$residuals, ylab=NA, xlab="cas")

m1<-tslm(log(GDP_train)~trend+I(trend^2))
summary(m1)
checkresiduals(m1)

```

```
plot(m1$residuals, ylab=NA, xlab="cas")
AIC(m2)
BIC(m3)

a<-accuracy(forecast(m5,h=12),log(GDP_test))
tab_df(as.data.frame(a),digits=4)

plot(forecast(m5,h=12), xlim=c(2010,2020),ylim=c(9.5,10),main="")
lines(log(GDP_test), col=2, lwd=2)
```